

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
17 January 2002 (17.01.2002)

PCT

(10) International Publication Number
WO 02/04496 A2

- (51) International Patent Classification⁷: **C07K 14/195**
- (21) International Application Number: PCT/US01/21525
- (22) International Filing Date: 6 July 2001 (06.07.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/216,750 7 July 2000 (07.07.2000) US
- (71) Applicant: **MEDIMMUNE, INC.** [US/US]; 35 West Watkins Mill Road, Gaithersburg, MD 20878 (US).
- (72) Inventors: **LANGERMANN, Solomon**; 6606 Cross Country Boulevard, Baltimore, MD 20878 (US). **REVEL, Andrew**; 5106 West Mockingbird Lane, Dallas, TX 75209 (US). **AUGUSTE, Christine**; 20147 Laurel Hill Way, Germantown, MD 20874 (US). **BURLEIN, Jeanne**; 5313 Juxon Place, Springfield, VA 22151 (US).
- (74) Agents: **GRANT, Alan, J. et al.**; Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart & Olstein, 6 Becker Farm Road, Roseland, NJ 07068 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:**
- without international search report and to be republished upon receipt of that report
 - with sequence listing part of description published separately in electronic form and available upon request from the International Bureau
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



WO 02/04496 A2

(54) Title: FIMH ADHESIN PROTEINS AND METHODS OF USE

(57) Abstract: The present invention provides bacterial immunogenic agents for administration to humans and non-human animals to stimulate an immune response. Also provided are methods for vaccination of mammalian species, especially human patients, with variants of the *E. coli* FimH protein, said variants being derived from different strains of *E. coli*, and to production of antibodies that protect the vaccine recipient against infection by pathogenic bacterial species. In another aspect the invention provides antibodies against such proteins and protein complexes that may be used as diagnostics and/or as protective/treatment agents for pathogenic bacterial species. A plasmid-based method of producing polypeptides, especially fused polypeptides, such as the complex of a bacterial chaperone and a bacterial adhesin, is also disclosed.

FIMH ADHESIN PROTEINS AND METHODS OF USE

5

This application claims the benefit of U.S. provisional application 60/216,750, filed 7 July 2000, the disclosure of which is hereby incorporated by
10 reference in its entirety.

FIELD OF THE INVENTION

15

This invention relates to bacterial adhesin proteins and active fragments thereof for use in vaccine compositions for the prevention, diagnosis and treatment of bacterial induced diseases such as those of the urinary tract, especially to the use of such adhesins as immunogenic agents in humans and
20 animals to stimulate an immune response.

BACKGROUND OF THE INVENTION

25

Urinary tract infections (herein, "UTI") present a disease process that is mediated by the attachment of bacteria to cells. *Escherichia coli* is the most common pathogen of the urinary tract, accounting for more than 85% of cases of asymptomatic bacteriuria, acute cystitis and acute pyelonephritis, as well as greater than 60% of recurrent cystitis, and at least 35% of recurrent
30 pyelonephritis infections. Furthermore, approximately 25%-30% of women experience a recurrent *E. coli* urinary tract infection within the first 12 months following an initial infection but after a second or third infection the rate of

recurrence increases to 60%-75%. Given the high incidence, continued persistence, and significant expense associated with *E. coli* urinary tract infections, there is a need for a prophylactic vaccine to reduce susceptibility to this disease.

5

While many factors contribute to the acquisition and progression of *E. coli* urinary tract infections, it is generally accepted that colonization of the urinary epithelium is a required step in the infection process. In a typical course of *E. coli* urinary tract infection, bacteria originate from the bowel, ascend into the bladder, and adhere to the bladder mucosa where they multiply and establish an infection (cystitis) before ascending into the ureters and kidneys. Disruption or prevention of pilus-mediated attachment of *E. coli* to urinary epithelia may prevent or retard the development of urinary tract infections. In this regard, a number of studies have pointed to a role for pili in mediating attachment to host uroepithelial cells.

10
15

To initiate infection bacterial pathogens must first be able to colonize an appropriate target tissue of the host. For many pathogens this tissue is located at a mucosal surface. Colonization begins with the attachment of the bacterium to receptors expressed by cells forming the lining of the mucosa. Attachment is mediated via proteins on the bacterium that bind specifically to cellular receptors. These proteins, or adhesins, are expressed either directly on the surface of the bacterium, or more typically, as components of elongated rod-like protein structures called pili, fimbriae or fibrillae.

20
25

Type 1 pili are thought to be important in initiating colonization of the bladder and inducing cystitis, whereas P pili are thought to play a role in ascending infections and the ensuing pyelonephritis.

Such pili are heteropolymeric structures that are composed of several different structural proteins required for pilus assembly. Two types of pili are of particular interest: P pili and type 1 pili. P pili-carrying bacteria recognize and

30

bind to the gal-(α 1-4)gal moiety present in the globoseries of glycolipids on kidney cells in mammals. Type 1 pili-carrying bacteria recognize and bind to D-mannose in glycolipids and glycoproteins of bladder epithelial cells.

- 5 FimH is the D-mannose-binding adhesin that promotes attachment of type 1 pillated bacteria to host cells via mannose-containing glycoproteins on eukaryotic cell surfaces. FimC is its periplasmic chaperone protein.

10 In this specification, the terms "pili", "fimbriae," and "fibrillae" are used to refer to heteropolymeric protein structures located on the extracellular surface of bacteria, most commonly gram-negative bacteria. Typically these structures are anchored in the outer membrane. Throughout this specification the terms pilus, pili, fimbriae, and fibrilla will be used interchangeably.

- 15 As used herein, the term "periplasmic chaperone" is defined as a protein localized in the periplasm of bacteria that is capable of forming complexes with a variety of chaperone-binding proteins via recognition of a common binding epitope (or epitopes). Chaperones serve as templates upon which proteins exported from the bacterial cell into the periplasm fold into their native
20 conformations. Association of the chaperone-binding protein with the chaperone also serves to protect the binding proteins from degradation by proteases localized within the periplasm, increases their solubility in aqueous solution, and leads to their sequentially correct incorporation into an assembling pilus.

- 25 Chaperone proteins are a class of proteins in gram-negative bacteria that are involved in the assembly of pili by mediating such assembly, but are not incorporated into the structure. FimC is the periplasmic chaperone protein that mediates assembly of type 1 pili in bacteria.

- 30 It has recently been reported that such chaperones can direct formation of the appropriate native structure of the corresponding adhesin or pilin by inserting a specific fold of the chaperone protein in place of a missing domain or

helical strand of the chaperone or pilin. Thus, FimH proteins tend to have their native structure in the presence of such a chaperone but not in its absence. [see: Choudhury et al, X-ray Structure of the FimC-FimH Chaperone-Adhesin Complex from Uropathogenic *E. coli*, *Science* 285, 1061 (1999); Sauer et al, 5 Structural Basis of Chaperone Function and Pilus Biogenesis, *Science* 285, 1058 (1999)] In addition, recent publications have indicated that the required chaperone strand can be inserted into the adhesin or pilin protein, such as FimH, to provide the missing structure and produce the correct native structure; Bamhart, M.M. et al., PapD-like Chaperones Provide the Missing Information for 10 Folding of Pilin Proteins, *Proc. Natl. Acad. Sci. (USA)*, 10.1073/pnas.130183897 (published online June 20, 2000).

Antibodies directed against purified whole type 1 or P pili protect against cystitis and pyelonephritis, respectively, in both murine and primate models for 15 these diseases. See: Abraham et al., *Infect Immun.* 48:625 (1985), Roberts et al., *Proc. Natl. Acad. Sci. (USA)* 91:11889 (1994), O'Hantey et al., *J. Clin. Invest.* 75: 347 (1985). However, such protection is limited to either homologous *E. coli* strains from which the pili used as immunogens were derived, or to a small subset of serologically cross-reactive heterologous strains. Therefore, vaccines 20 composed predominantly of the major structural proteins of pili (i.e., PapA or FimA) appear to be of limited value because antibodies developed against these highly variable proteins are specific for the strains used for immunization.

Furthermore, antibodies to FimH have been found to be protective. 25 Bamhart, M.M. et al., PapD-like Chaperones Provide the Missing Information for Folding of Pilin Proteins, *Proc. Natl. Acad. Sci. (USA)*, 10.1073/pnas.130183897 (published online June 20, 2000).

Recently, Sokurenko et al [see: *J. Bacteriol.* 177, 3680-86 (1995)] have 30 found that quantitative variations in mannose-sensitive adhesion of *E. coli* are due primarily to structural differences in the FimH adhesin. Further research has shown that the ability of the FimH lectins to interact with monomannosyl

residues strongly correlates with their ability to mediate *E. coli* adhesion to uroepithelial cells so that certain phenotypic variants of type 1 fimbriae may contribute more than others to the virulence of *E. coli* in the urinary tract. [Sokurenko et al, *J. Biol. Chem.* **272**, 17880-17886 (1997)]. Heretofore, random
5 point mutations in FimH genes that increase binding of the adhesin to monomannose residues (structures abundant in the oligosaccharide moieties of urothelial glycoproteins) have been found to confer increased virulence in the mouse urinary tract. [See: Sokurenko et al, *Proc. Natl. Acad. Sci. USA* **95**, 8922-8926 (1998)]

10

BRIEF SUMMARY OF THE INVENTION

In one aspect, the present invention relates to immunogenic
15 polypeptides derived from different strains of the bacterium *Escherichia coli* (*E. coli*) which differ from each other in one or more amino acid residues and that induce immunological responses that lead to protection of an animal, especially a human patient, following vaccination with said polypeptides.

20 It is an object of the present invention to provide immunogenic FimH polypeptides useful as components of vaccines for the prevention and treatment of infections caused by *E. coli*, in particular, urinary tract infections. In specific embodiments, the polypeptides of the present invention have the amino acid sequences shown in Figure 2. Such vaccine compositions comprising the novel
25 polypeptides disclosed herein are useful in vaccination and treatment of urinary tract infections, especially those caused by *E. coli*. In one embodiment, vaccine compositions according to the invention comprise the polypeptides of Figure 2 (or truncated segments of the sequences of Figure 2 which truncates have mannose-binding ability and can serve as vaccines), as stabilized structures in
30 the form of complexes with a chaperone, such as FimC, or with a portion of the sequence of FimC which portion serves to stabilize the structure of the FimH

adhesin. A particular embodiment employs the consensus sequence of Figure 2 (SEQ ID NO: 55).

It is another object of the present invention to provide antibodies, including both polyclonal and monoclonal antibodies, with specificity for the novel polypeptides disclosed herein and whose amino acid sequences are shown in Figure 2.

It is a still further object of the present invention to provide methods of prophylaxis for the prevention of urinary tract infections using the vaccine compositions disclosed herein.

It is a yet still further object of the present invention to provide methods of treatment of diseases of the urinary tract comprising the use of vaccine compositions as disclosed herein and the use of antibodies generated against such vaccine compositions, and the polypeptides contained therein.

It is another object of the present invention to provide a novel method of preparing polypeptides from recombinant cells using a vector comprising the plasmid of Figures 3 through 6. In specific embodiments, the polypeptides comprise the amino acid sequences of Figure 2. In one embodiment, this process and plasmid can be used to prepare polypeptides that comprise a bacterial chaperone, such as FimC, fused to a bacterial adhesin, such as FimH or any of the polypeptides presented in Figure 2, including any consensus sequence derived therefrom.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the nucleotide sequence encoding a number of FimH polypeptides disclosed according to the present invention and derived from *E. coli*. Figures 1(a) through 1(j) follow in sequence.

Figure 2 shows the amino acid sequences, and sequence variations, for a number of FimH proteins of different strains of *E. coli* with the particular strain listed at the left. Figures 1(a) through 1(c) follow in sequence. A consensus sequence is presented at the bottom. The sequence for FimH of J96 is known in the art.

Figure 3 is a diagram showing the construction of the vector designated as pCGA101-8.

Figure 4 is a diagram showing the construction of the vector designated as pCGA122-30.

Figure 5 is a diagram showing the steps in selection of the final clone pCGA139-1-1.

Figure 6 is a diagram showing the arrangement of genes in the vector designated as pCGA139-1-1.

DETAILED DESCRIPTION OF THE INVENTION

It is an object of the present invention to utilize an immunogenic composition for a vaccine (or to produce antibodies for use as a diagnostic or as a passive vaccine) comprising novel polypeptides, in particular, FimH polypeptides or variants thereof, either alone or in complexed form with a chaperone, or stabilization-inducing portion of a chaperone, or as a series of isolated non-contiguous domains, especially mannose-binding domains, linked together to form a polypeptide, or polypeptide-like, structure.

In one embodiment of the present invention, FimH proteins (naturally or recombinantly produced, as well as functional analogs) from bacteria that

produce type 1 pili are contemplated. Even more particularly, *E. coli* FimH proteins are contemplated.

The present invention provides isolated polypeptides comprising a
5 sequence selected from the group consisting of at least residues 26 to 186 of
SEQ ID NO: 23 through 45 and 55.

The present invention further relates to isolated polypeptides further
comprising about the N-terminal two thirds (which includes about residues 1
10 through 186) of the sequences selected from SEQ ID NO: 23 through 45 and
55.

The present invention still further relates to isolated polypeptides
comprising an amino acid sequence selected from the group consisting of
15 SEQ ID NO: 23 through 45 and 55.

The polypeptides of the present invention may conveniently be present
in the form of a composition comprising one or more of the polypeptides in
any desired combination or relative concentrations. Further, the present
20 invention includes active truncates, portions, fragments or segments of the
novel polypeptides disclosed herein.

The present invention also relates to an isolated polynucleotide
encoding one or more of the polypeptides of the present invention, or active
25 truncates, fragments, portions or segments thereof. In one embodiment, the
isolated polynucleotides are shown in Figure 1(a) – 1(j).

A vaccine composition according to the present invention is one
comprising an immunogenically effective amount of a polypeptide of the
30 invention, including immunogenically active truncates, portions, fragments and
segments thereof, and in any and all active combinations thereof, wherein
said polypeptide, or active fragment, or fragments, is/are suspended in a

pharmacologically acceptable carrier, which includes all suitable diluents or excipients.

5 The present invention also relates to antibodies, either polyclonal or monoclonal, either recombinant or synthetic, specific for any of the polypeptides, as disclosed herein (SEQ ID NO: 23 through 45 and 55) including immunogenically active truncates, portions, fragments and segments of such polypeptides, which antibodies may be used alone or in any and all active combinations thereof.

10

Such polypeptides, where the entire sequence is used to form a vaccine, are commonly structurally stabilized by complexing with a chaperone, especially FimC (so as to form a FimCH complex – see: Choudhury et al (1999) and Sauer et al (1999), the disclosures of which
15 papers are hereby incorporated by reference in their entirety) or where such structural stabilization is produced by complementing the structure of the adhesin with a portion of a chaperone, such as FimC, that induces the stabilization of the adhesin for use as a vaccine [see Barnhart et al. supra, the disclosure of which is hereby incorporated by reference in its entirety), or
20 where the mannose-binding portions of the sequences of the adhesins of Figure 2 are isolated from the native molecules, or synthesized, and then joined together to form a structure comprising 1, 2, 3 or more such domains as part of a single polypeptide (other than the native protein) or where such domains are held together by some other chemical linkages.

25

Because of the immunogenicity and protective properties of the polypeptides and fragments disclosed herein, the present invention further relates to methods for preventing and/or treating disease in an animal, especially a human patient afflicted therewith, comprising administering to
30 said animal an amount of a vaccine composition of the present invention that is effective for such prevention and/or treatment. In specific embodiments of the present invention, said disease is a urinary tract infection or a bladder

infection. In preferred embodiments, said disease is caused by a bacterium of the family Enterobacteriaceae, especially *E. coli*, that is effective for such preventing and/or treating.

- 5 The present invention further relates to methods of treating disease in an animal, especially a human patient afflicted therewith, comprising administering to said animal a therapeutically effective amount of an antibody of the invention as disclosed hereinabove. In specific embodiments of the present invention, said disease is a urinary tract infection or a bladder
10 infection. In preferred embodiments, said disease is caused by a bacterium of the family Enterobacteriaceae, especially *E. coli*.

- The present invention also relates to a recombinant cell expressing a polypeptide, and fragments thereof, selected from the group consisting of
15 SEQ ID NO: 23 through 45, and 55. Here, the term "expressing" means synthesizing a polypeptide from the disclosed polynucleotides of the invention.

- The present invention also relates to a vector comprising a
20 polynucleotide selected from the polynucleotides encoding the polypeptides, and active fragments thereof, of the invention.

- In accordance with the disclosure herein, the present invention also includes an immunogenic complex comprising a periplasmic chaperone and a
25 polypeptide, or active fragment thereof, selected from the group consisting of the polypeptides of SEQ ID NO: 23 through 45, and 55.

- A preferred embodiment of such an immunogenic composition is a composition wherein the active component of said immunogenic composition is
30 a member selected from mannose-binding fragments of FimH adhesin protein variants, or where said proteins are donor complemented single polypeptide structures or present in the form of an immunogenic complex comprising non-

contiguous mannose-binding domains derived from said variants (see Langermann, U.S. Patent Application Serial No. 60/144,016, filed July 15, 1999, the disclosure of which is hereby incorporated by reference in its entirety).

5 In another aspect of the invention, immunogenic compositions of the invention may be utilized to produce antibodies to diagnose or treat diseases, especially urinary tract infections, or to produce vaccines for prophylaxis and/or treatment of such infections as well as booster vaccines to maintain a high titer of antibodies against the immunogen(s) of the immunogenic composition.

10

Applicant has found that FimH polypeptides are highly conserved between various strains of *E.coli*. Moreover, the amino acid changes that occur between strains generally occur at similar amino acid positions.

15

As a result of the high conservation of FimH between *E.coli* strains, FimH polypeptides from one strain are capable of inducing antibody responses that inhibit or prevent other *E.coli* strains from binding to cells by a FimH lectin and/or provide protection and/or treatment against infection caused by other *E.coli* strains. The SEQ ID NOs. for the FimH proteins (including a consensus sequence) and polynucleotides of the present invention are provided in Table 1 and correlate with the strain designations provided in Figures 1 and 2, and the data of Table 2.

20

25 There are various ways in which the polypeptides disclosed according to the present invention can be utilized to form therapeutically effective vaccine compositions for use in the prevention and/or treatment of urinary tract and bladder infections, especially those caused by *E. coli*.

30 In accordance with the teachings herein, the present invention is directed to an immunogenic composition comprising a purified complex of a periplasmic chaperone protein, especially FimC, and a FimH variant according to the invention. The FimC may be obtained from the same *E.coli* strain as the FimH

or from a different strain. The FimH polypeptide, or portion thereof, is maintained in the complex in an immunogenic form capable of inducing an immune response when appropriately introduced into a human patient in need thereof. As used herein, the term "patient in need thereof" refers to a human that
5 is infected with, or at risk of being infected with, pathogenic bacteria that produce pili, especially *E. coli* and related bacteria. For research purposes, a mouse model can be utilized to simulate such a patient in some circumstances. See Choudhury, et al., supra, and Sauer, et al., supra.

10 The polypeptides according to the present invention can also be utilized as components of therapeutically effective vaccine compositions by forming the native structures of said FimH variants without the need for a complex with FimC itself. Thus, the FimH variants as disclosed herein are readily produced by recombinant methods in such a way as to incorporate therein the sequence of
15 FimC required for stabilization (see Barnhart et al (2000)).

An additional means of utilizing the novel polypeptides, or FimH variants, of the present invention is to form synthetic structures comprising non-contiguous domains of said variants. It is known that the cell binding portions of
20 FimH are generally composed of the mannose-binding segments, formed of about the N-terminal two thirds of the molecule. The remaining pilin-binding portion is the segment that interacts with FimC to form a complex in the fibrillum of the bacterial cell. Thus, the FimH variants of the present invention are readily engineered to produce only the specific, and relatively short, mannose-binding
25 domains of the N-terminal two thirds of the sequences depicted in Figure 2. All or a portion of these domains may be strung together using convenient linker sequences, or other linking structures, to provide polypeptides composed of non-contiguous mannose binding domains, the overall structure of which provides a highly immunogenic structure for use in the vaccine compositions
30 disclosed herein (see Langermann supra).

Of course, these are only three means of utilizing the FimH variants disclosed according to the present invention and other useful embodiments will suggest themselves to those skilled in the relevant arts from the teachings herein.

5

The proteins, and immunologically active fragments thereof, of the present invention are useful immunogens for preparing vaccine compositions that stimulate the production of antibodies that confer immunity to pathogenic species of bacteria. Further, preparation of vaccines containing purified proteins
10 as antigenic ingredients is well within the level of skill in the art.

The pharmaceutical compositions useful herein also contain a pharmaceutically acceptable carrier, including any suitable diluent or excipient, which includes any pharmaceutical agent that does not itself induce
15 the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Pharmaceutically acceptable carriers include, but are not limited to, liquids such as water, saline, glycerol and ethanol, and the like, including carriers useful in forming sprays for nasal and other respiratory tract delivery or for
20 delivery to the ophthalmic system.

Vaccine compositions may, and usually do, incorporate additional substances to stabilize pH, or to function as adjuvants, wetting agents, or emulsifying agents, which can serve to improve the effectiveness of the vaccine.

25

Vaccines are generally formulated for parenteral administration and are injected either subcutaneously or intramuscularly. Such vaccines can also be formulated as suppositories or for oral administration, using methods known in the art.

30

The amount of vaccine sufficient to confer immunity to pathogenic bacteria is determined by methods well known to those skilled in the art. This

quantity will be determined based upon the characteristics of the vaccine recipient, including considerations of age, sex, and general physical condition, and the level of immunity required. Where vaccines are administered by subcutaneous or intramuscular injection, a range of 1 to 500 µg purified protein
5 may be given.

In addition to use as vaccines, the FimH variants disclosed herein are available for use as immunogens to stimulate the production of antibodies for use in passive immunotherapy, for use as diagnostic reagents, and for use as
10 reagents in other processes such as affinity chromatography.

In one aspect of the invention complexes comprising the *E. coli* chaperone FimC and a FimH variant of the invention may be formed by co-expressing the appropriate FimH variant polypeptide (of a sequence according to Figure 2) along with FimC, whose amino acid and nucleotide sequences are
15 known in the art, from a recombinant cell.

In addition, the FimC-FimH complexes useful in vaccines can be recovered from the periplasmic spaces of cells of the indicated strains disclosed herein. These complexes are found in relatively large amounts in recombinant
20 *E. coli* strains which express the FimC protein at levels in excess of those produced in wild type strains. A suitable recombinant strain is C600/pHJ9205, a strain in which expression of FimC has been put under control of the arabinose promoter. Those skilled in the art will recognize that other promoter sequences that can be regulated easily may also be used. Of course, such cells are readily
25 engineered to express one or more of the FimH variant polypeptides of the invention. An extract of periplasm is obtained by exposing the bacteria to lysozyme in the presence of a hypertonic sucrose solution. FimC-H complexes can also be purified using conventional protein purification methods well known in the art.

30

In a similar manner, FimH fragments that are recombinantly produced either by having *E. coli* produce the full-length FimH variant polypeptides of the

present invention and then fragmenting the protein, or the fragment itself, are produced recombinantly -may be isolated by mannose-binding affinity purification. Thus, only fragments of the FimH protein that retain mannose binding are isolated. Preferably, such mannose-binding fragments have a label
5 such as a his-tag included and may be purified by methods such as Nickel chromatography.

As used herein, the terms "portion," "segment," and "fragment," when used in relation to polypeptides, refer to a continuous sequence of residues, such as amino acid residues, which sequence forms a subset of a larger
10 sequence. For example, if a polypeptide were subjected to treatment with any of the common endopeptidases, such as trypsin or chymotrypsin, the oligopeptides resulting from such treatment would represent portions, segments or fragments of the starting polypeptide. When used in relation to a polynucleotides, such
15 terms refer to the products produced by treatment of said polynucleotides with any of the common endonucleases.

The polynucleotides encoding the polypeptides of the invention, for example, those of Figure 1, may have the coding sequence fused in frame to a marker sequence which allows for purification of the polypeptides of the present invention. The marker sequence may be, for example, a hexa-histidine tag supplied by a pQE-9 vector to provide for purification of the mature polypeptides fused to the marker in the case of a bacterial host, or, for example, the marker
20 sequence may be a hemagglutinin (HA) tag when a mammalian host, e.g. COS-7 cells, is used. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein (Wilson, I., et al., Cell, 37:767 (1984)).

The present invention also relates to vectors which include
30 polynucleotides encoding one or more of the adhesin proteins of the present invention, host cells which are genetically engineered with vectors of the invention and the production of such adhesin proteins and/or chaperone

proteins by recombinant techniques in an isolate and substantially immunogenically pure form.

Host cells are genetically engineered (transduced or transformed or
5 transfected) with the vectors comprising a polynucleotide encoding a chaperone, adhesin protein, mannose binding fragment of an adhesin protein, or the like of this invention which may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient
10 media modified as appropriate for activating promoters, selecting transformants or amplifying the polynucleotides which encode such polypeptides. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

15

Vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and
20 pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host. Such procedures and others are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an
25 appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter, the *E. coli. lac* or *trp*, the phage lambda P_L promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also
30 contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin or kanamycin resistance in E. coli.

For example, optimal expression of a FimH-C complex has been achieved using a newly constructed single vector containing the FimH and FimC genes but having the advantage that each gene is under its own separate lac promoter. Thus, one lac promoter is 5' with respect to FimC while the second lac promoter is 5' to the FimH gene. This plasmid was successfully constructed using the common plasmid pUC19 as a background vector [Yannish-Perron, C., Vierira, J. and Messing, J., *Gene*, 33:103-119 (1985)]. This new plasmid, when used to transform the host *E. coli* strain BL21 [as described in Phillips, T.A., Van Bogelen, R. A., and Neidhart, F.C., *J. Bacteriol.* 159:283-287 (1984)] and then induced using IPTG at the mid-logarithmic stage of growth, gives maximal expression of the FimCH complex in the bacterial periplasmic space. This material is then extracted and purified by methods well known in the art, including those described herein.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the proteins.

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as *E. coli*, *Streptomyces*, *Salmonella typhimurium*; fungal cells, such as yeast; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9; animal cells such as CHO, COS or Bowes melanoma; adenoviruses; plant cells, etc. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

More particularly, the present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example. Bacterial: pQE70, pQE60, pQE-9 (Qiagen, Inc.), pbs, pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia). However, any other plasmid or vector may be used as long as they are replicable and viable in the host.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P_R, P_L and TRP. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In a further embodiment, the present invention relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis, L., Dibner, M., Battey, J., Basic Methods in Molecular Biology, (1986)).

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the polypeptides of the invention can be synthetically produced by conventional peptide synthesizers.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts, as well as other methods in molecular biology, are described in Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor, N.Y., (1989), Wu et al, *Methods in Gene Biotechnology* (CRC Press, New York, NY, 1997), and *Recombinant Gene Expression Protocols*, in *Methods in Molecular Biology*, Vol. 62, (Tuan, ed., Humana Press, Totowa, NJ, 1997), the disclosures of which are hereby incorporated by reference.

Transcription of the DNA encoding the polypeptides of the present invention by higher eukaryotes is increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act on a promoter to increase its transcription. Examples including the SV40 enhancer on the late side of the replication origin bp 100 to 270, a cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), α -

factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification
5 peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable
10 translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and
15 various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of
20 replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter
25 and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are
30 cultured for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by
5 any convenient method, including freeze-thaw cycling, sonication, a french press, mechanical disruption, or use of cell lysing agents, such methods are well know to those skilled in the art.

Various mammalian cell culture systems can also be employed to
10 express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell, 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and
15 enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

20

The polypeptides can be recovered and/or purified from recombinant cell cultures by well-known protein recovery and purification methods. Such methodology may include ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose
25 chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. In this respect, chaperones may be used in such a refolding procedure. Finally, high performance liquid chromatography (HPLC)
30 can be employed for final purification steps.

The polypeptides that are useful as immunogens in the present invention may be a naturally purified product, or a product of chemical synthetic procedures, or produced by recombinant techniques from a prokaryotic or eukaryotic host (for example, by bacterial, yeast, higher plant, insect and mammalian cells in culture). Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated.

Procedures for the isolation of a periplasmic chaperone protein complexed with an adhesin protein are known in the art, as an example see Jones et al., *Proc. Natl. Acad. Sci. (USA)* 90:8397-8401 (1993). Further, the individually expressed adhesin proteins may be isolated by recombinant expression/isolation methods that are well-known in the art. Typical examples for such isolation may utilize an antibody to the protein or to a His tag or cleavable leader or tail that is expressing as part of the protein structure.

It is contemplated that the polypeptides of the present invention may be in isolated or purified form.

"Isolated" in the context of the present invention with respect to polypeptides (or polynucleotides) means that the material is removed from its original environment (e.g., the cells used to recombinantly produce the polypeptides disclosed herein). Such peptides could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment. The polypeptides and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The recombinant and/or immunogenic polypeptides, disclosed in accordance with the present invention, may also be in "purified" form. The term "purified" does not require absolute purity; rather, it is intended as a relative definition, and can include preparations that are highly purified or

preparations that are only partially purified, as those terms are understood by those of skill in the relevant art. For example, polypeptides from individual clones isolated from a cDNA library have been conventionally purified to electrophoretic homogeneity. Purification of starting material or natural
5 material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. Furthermore, claimed polypeptides having a purity of preferably 0.001%, or at least 0.01% or 0.1%, and even desirably 1% by weight or greater is expressly contemplated.

10

For purposes of recombinantly producing the polypeptides of the invention, the term "expression product" means that polypeptide or protein that is the natural translation product of the gene and any nucleic acid sequence coding equivalents resulting from genetic code degeneracy and
15 thus coding for the same amino acid(s).

The FimCH polypeptides useful in forming the vaccine compositions of the present invention may conveniently be cloned using various cloning systems. An example of a useful cloning system for synthesizing FimCH
20 molecules that may incorporate any of the FimH sequences disclosed herein (see Figure 2) is presented in Example 2 and utilizes a plasmid based cloning system. The FimCH complex described therein is composed of a 52 kDa complex composed of two proteins: FimC (22.8 kDa) and FimH (29.1 kDa) in a 1:1 equimolar ratio. The FimCH complex is expressed from a pUC-based
25 vector (pGCA139-1-1) with two separate lac-inducible promoters driving expression of the FimC and FimH genes, respectively. The FimC and the FimH genes in the pGCA139-1-1 vector were derived from uropathogenic *E. coli* isolate J96.

30 The FimCH complex is produced in the periplasm of *E. coli* strain BL21 and is purified from periplasmic extracts by standard chromatographic methods. The FimCH protein has been formulated in a number of different

buffers compatible with its solubility profile including 20 mM HEPES (pH 7.0), PBS (pH 7.0) and sodium citrate (pH 6.0) + 0.2 M NaCl. This sodium citrate/sodium chloride formulation enhances the stability of the FimCH complex and is also compatible with commonly used diluents.

5

Plasmid pCGA139-1-1 was constructed as a means of producing relatively large amounts of *E. coli* chaperone-adhesin complex, FimCH, for use in the vaccine compositions disclosed herein. Such vaccines act by blocking adherence of *E. coli* to the mucosa of the urinary tract thus preventing colonization which often results in infection.

10

The plasmid vector, pCGA139-1-1, contains the following genetic elements: (1) an *E. coli* *FimC* chaperone gene followed by (2) the *fimH* adhesin gene, both from *E. coli* strain J96 (a urinary tract infection (UTI) isolate) each preceded by its respective native signal sequence (*nss*); (3) a kanamycin resistance (*kan^r* or *k^r*) marker; (4) *lac^r* which codes for a repressor protein that binds the *lac* promoter unless it is induced; (5) an inactivated beta-lactamase (*bla*) gene; (6) pUC origin of replication (*ori*); and (7) two *lac* promoters, one preceding the *fimC* signal and the other preceding that of *fimH*.

15

20

The polypeptides, their fragments or other derivatives, or analogs thereof, or cells expressing them can be used as an immunogen to produce antibodies thereto. These antibodies can be, for example, polyclonal or monoclonal antibodies. The present invention also includes chimeric, single chain, and humanized antibodies, as well as Fab fragments, or the product of an Fab expression library. Various procedures known in the art may be used for the production of such antibodies and fragments.

25

Antibodies generated against the variant FimH polypeptides corresponding to a sequence of the present invention can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides

30

to an animal, preferably a nonhuman. The antibody so obtained will then bind the polypeptides itself. In this manner, even a sequence encoding only a fragment of the polypeptides can be used to generate antibodies binding the whole native polypeptides.

5

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, *Nature*, 256:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., *Immunology Today* 4:72 (1983)), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole, et al. (1985) in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to immunogenic polypeptide products of this invention. Also, transgenic mice may be used to express humanized antibodies to immunogenic polypeptide products of this invention.

In order to facilitate understanding of the above description and the examples which follow below certain frequently occurring methods and/or terms will be described.

"Plasmids" are designated by a lower case p preceded and/or followed by capital letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord with published procedures. In addition, equivalent plasmids to those described are known in the art and will be apparent to the ordinarily skilled artisan.

30

In carrying out the procedures of the present invention it is of course to be understood that reference to particular buffers, media, reagents, cells,

culture conditions and the like are not intended to be limiting, but are to be read so as to include all related materials that one of ordinary skill in the art would recognize as being of interest or value in the particular context in which that discussion is presented. For example, it is often possible to substitute one
5 buffer system or culture medium for another and still achieve similar, if not identical, results. Those of skill in the art will have sufficient knowledge of such systems and methodologies so as to be able, without undue experimentation, to make such substitutions as will optimally serve their purposes in using the methods and procedures disclosed herein.

10 The present invention will now be further described by way of the following non-limiting examples. In applying the disclosure of these examples, it should be kept clearly in mind that other and different embodiments of the methods disclosed according to the present invention will no doubt suggest
15 themselves to those of skill in the relevant art.

EXAMPLE 1

20 Passive immunization using the FimH variants of the present invention was demonstrated as follows. Anti-serum against FimC and FimCH complexes (using strain J96 (SEQ ID NO: 44) as the source of FimH and strain NU14 as source of FimC) was generated as two different pools and used for these experiments. Mice were passively immunized intraperitoneally with 100 μ l each
25 of either anti-C or anti-FimCH rabbit sera 24 hours and 4 hours prior to inoculation. Endpoint titers for the rabbit sera were determined to be at least 1:500,000 by ELISA against the respective antigens.

Bacteria of the respective strains were then collected, washed and re-
30 suspended in phosphate buffered saline (PBS) and cell concentration adjusted to OD = 1.8 (at 600 nm). This suspension was then diluted 1:10 in PBS and tested for hemagglutination (HA) with guinea pig erythrocytes. This final suspension was used as inoculum and viability was determined on TSA plates.

- Mice were anaesthetized and then inoculated intraurethrally with 50 μ l of *E. coli* suspension containing about 3×10^7 cfu (colony forming units). Two days post-inoculation, the mice were sacrificed and bladders were removed and collected into 500 μ l PBS supplemented with 1% mannose. The number of CFU's per
- 5 bladder was determined by grinding the bladders with a tissue tearer and then diluting and plating the suspension on TSA plates. The mean number of colony forming units per bladder was determined and data transformed to log CFU/bladder (as reported in Table 2).

Table 1. SEQ ID NOs. of Nucleotides and Amino acids

| Strain | Nucleotide SEQ ID NO: | Polypeptide SEQ ID NO: |
|-----------|--------------------------|---------------------------|
| 5 | | |
| J96 | 21 | 44 |
| NU14 | 22 | 45 |
| B210 | 1 | 23 |
| B212 | 2 | 24 |
| 10 B217 | 3 | 25 |
| B223 | 4 | 26 |
| B228 | 5 | 27 |
| B238 | 6 | 28 |
| B240 | 7 | 29 |
| 15 B242 | 8 | 30 |
| DS17 | 9 | 31 |
| EC42 | 10 | 32 |
| EC45 | 11 | 33 |
| EC56 | 12 | 34 |
| 20 EC58 | 13 | 35 |
| EC60 | 14 | 36 |
| EC61 | 15 | 37 |
| EC62 | 16 | 38 |
| EC80 | 17 | 39 |
| 25 EC89 | 18 | 40 |
| EC95 | 19 | 41 |
| G162 | 54 | 42 |
| G189 | 20 | 43 |
| Consensus | — | 55 |
| 30 | | |

Table 2. Passive Protection by FimH Variants

| Strain | Mean Log CFU per Bladder | | | T- test | |
|--------|--------------------------|-------|-------|-----------|--------------|
| | FimC | FimCH | Naïve | C vs. CH | CH vs. Naïve |
| B223 | 7.79 | 5.69 | 7.58 | 0.0034 | 0.0107 |
| EC45 | 6.43 | 4.58 | ND | 0.0087 | ND |
| Nu14 | 4.54 | 2.53 | 5.22 | 0.0014 | 0.0000428 |
| B217 | 4.47 | 3.49 | 5.17 | 0.0142 | 0.0007 |
| DS17 | 4.64 | 3.02 | 4.45 | 0.0163 | 0.0355 |
| B218 | 4.30 | 2.99 | 4.16 | 0.0066 | 0.0331 |
| B220 | 4.18 | 1.93 | 3.55 | 0.0000257 | 0.0016 |
| EC56 | 3.02 | 2.60 | 3.34 | 0.5245 | 0.2222 |
| EC42 | 2.47 | 1.13 | 2.83 | 0.0274 | 0.0013 |
| J96 | 2.09 | 0.96 | 2.29 | 0.1005 | 0.0328 |
| B212 | 3.20 | 2.05 | 3.20 | 0.0167 | 0.443 |

5

EXAMPLE 2**Construction of Plasmid Vector for FimCH Production**

The plasmid vector, pCGA139-1-1, for production of FimCH was constructed in several steps. Although this example is directed to the use of FimH and FimC from J96, the example is also applicable to the FimH variants of the invention.

10

Construction of pCGA101-8

Genomic DNA was prepared from *Escherichia coli* strain J96. The pellet from 1.0 ml of an overnight culture was washed with PBS, resuspended in 500 μ l sterile sucrose Tris EDTA (0.3 molar sucrose: 100 millimolar Tris, pH8.0: 50 millimolar EDTA) and 0.01 μ g lysozyme was added thereto. The suspension was incubated at 37°C for 10 minutes and SDS was added to a final concentration of 0.5%. The mixture was then treated with RNase for 10 minutes at 37°C after

15

which the DNA was extracted with water saturated phenol followed by a mixture of chloroform: isoamyl alcohol at a 24:1 ratio v/v and ethanol precipitated. The resulting pellet was washed with 70% ethyl alcohol, dried and resuspended in a solution containing 10.0 mM Tris and 0.1 mM EDTA. This
5 DNA was used as template for PCR amplification of the *fimC* gene.

PCR was performed with primers ga1F and ga2R (SEQ ID NO: 56 and 57, respectively) containing NcoI and BglII restriction sites, respectively. Conditions used were: 1 cycle at 95°C for 1.0 min; 25 cycles of amplification with
10 strand separation at 95°C for 30 sec, annealing at 50°C for 30 sec and strand elongation at 72°C for 2.0 min. This was followed by one 10 min cycle at 72°C to ensure complete elongation of all ends. PCR products were purified via Qiagen columns and the gene was cloned into the vector pPW19R (construction described below), 3' to the Pel B leader sequence on the plasmid. The result
15 was plasmid pCGA101-8 (shown in Figure 3).

Construction of pCGA122-30

A kanamycin resistance gene was excised as a AlwNI/StyI fragment from pET26b(+) (Novagen) and cloned into the unique DralI site in pTTQ18 5' of the
20 *lac^R* producing pTTQ18K. This plasmid contains the *lac^R* and *kan^r* genes in tandem so they can be cloned as a single cassette (see Figure 4).

In an effort to obtain optimal yields of FimH, its native signal was used to replace the Pel B leader sequence. A method utilizing overlapping PCR was
25 used. Primary PCR segments were (1) *fimH* gene and its native signal from genomic J96 DNA with primers ga13F and ga6R (SEQ ID NO: 58 and 59, respectively) and (2) lac promoter/operator (lac p/o) from pPW19R with primers ga11F and ga9R (SEQ ID NO: 60 and 61, respectively). Overlapping PCR resulted in a single fragment primed with ga11F and ga6R containing BglII and
30 Sall sites, respectively.

Vector pCGA122-30 was made by ligating (1) the BglII/Sall PCR fragment consisting of *lac p/o*, *fimH* native signal sequence + *fimH*; (2) the BglII/Scal fragment from pCGA101-8 containing the *beta-lactamase* gene, pUC *ori*, *lac p/o*, *pelB* leader, and *fimC*; and (3) the cassette containing *lacI^d* and *kanI^r* from pTTQ18K as a Sall/Scal fragment.

Construction of pGA139-1 and selection of the final clone

The *PelB* signal 5' to *fimC* was replaced with *fimC* native signal: primary fragments for *fimC* native signal replacement were derived from (1) genomic J96 DNA with primers ga24F and ga23R, and (2) *lac p/o* from pPW19R with primers as mentioned previously. *FimC* and its native signal were the result of overlapping PCR obtained as one cassette with primers ga24F and ga2R containing AflIII and BglII sites, respectively. The product was cloned as a replacement AflIII/EcoRI to pCGA122-30 producing pCGA126-1 (see Figure 5 and SEQ ID NO: 46).

The *beta-lactamase (bla)* gene was inactivated by interruption at the Scal site. This was followed by treatment with an exonuclease (Bal31) and subsequent filling in with deoxynucleotide tri-phosphates (dNTPs). The plasmid was finally re-ligated resulting in a deletion of about 60 bases and thus forming the plasmid pCGA139-1 (see Figures 5 and 6).

pCGA126-1 was sequenced in its entirety and the deletion at the Scal site (thereby giving rise to the pCGA139-1 plasmid) was confirmed by sequencing.

Plasmid pCGA139-1 was transformed into a BL21 *E. coli* host strain to optimize protein expression. Twenty microliters of BL21 competent cells were pipetted into a pre-chilled 1.5 ml polypropylene tube. One microliter of vector was added and the mixture was incubated on ice for 5 min. The cells were heat shocked in a 42°C water bath for exactly 30 sec followed by incubation on ice for 2 min. SOC medium (80 µl) was added followed by incubation at 37°C with

shaking at 250 rpm for 1 hr. The culture was plated on 2XYT agar containing 50 µg/ml kanamycin and plates were incubated overnight at 37°C. The recipes for terrific medium and SOC broth can be found in standard books on cell culture (for example, see Sambrook et al (1989), *supra*, at book 3, appendix page A.2.).

5

Plasmid preparations and frozen 15-20% glycerol stocks were made from individual colonies grown overnight in terrific broth. Candidates were chosen by both the absence of the *Scal* site and by sensitivity to 50 µg/ml ampicillin. Plasmids were further analyzed for production of target protein. pCGA139-1 was
10 cloned by plating on 2XYT agar containing 50 µg/ml kanamycin. Six clones were analyzed according to their restriction patterns, Western blot analysis, and production of FimCH protein. A single colony was grown overnight in broth and stored in 15%-20% sterile glycerol in Nunc vials at a temperature of -70°C.

15

Overnight cultures were diluted 1:30 in terrific broth containing 50 µg/ml kanamycin and grown at 37°C to mid-log phase (about 0.3 at OD₆₀₀). Fifteen ml of each culture was induced with 2.0 mM IPTG and harvested after 3 hr. Several 1 ml aliquots from each sample were sedimented in an eppendorf centrifuge at 14,000 rpm for 2 min. Total protein was estimated by BCA assay and 1.0 µg
20 total protein of uninduced and induced culture was loaded to two polyacrylamide gels for electrophoresis against FimCH standards of known concentration. FimC and FimH were assayed on separate gels/membranes. Samples were also assayed via ion exchange chromatography for levels of FimCH protein.

25

Proteins were transferred to nitrocellulose membranes via Western blot, blocked with 2% dried milk and treated with primary polyclonal antibodies to FimC or FimH-T3. Membranes were washed 3X (15 min each wash) with PBS plus 0.01% Tween 20 after which a donkey anti-rabbit secondary antibody conjugated to horseradish peroxidase (HRP) was applied for 1 hr. Membranes
30 were washed as described before followed by treatment with an anti-HRP detection reagent, ECL, or ECL-plus. Nitrocellulose was finally exposed to x-ray films and developed in a M35-A x-omatic processor.

For ion exchange chromatography of the FimCH product, all samples were resuspended in 200 μ l of PBS, sonicated for 12 min, and diluted 4 fold with PBS. Each sample was centrifuged at 10,000 rpm for 3 min into a 0.45 micron spin filter unit and transferred into HPLC microvials for analysis. A Pharmacia Mono-S HR 5/5 column (5 mm x 50 mm) was used for the quantification of pilus proteins in analyzed samples. Mobile phase A was 20 mM potassium phosphate (pH 7.0); mobile phase B was mobile phase A but containing 0.5 M potassium chloride. A gradient of 0%-30% B over 20 minutes was run at a flow rate of 1.25 ml/min. Eluted protein was detected using intrinsic tryptophan fluorescence detection (excitation 280 nm, emission 335 nm). A standard curve was generated using reference standard material diluted to concentrations from 5.2 μ g/ml to 15.6 μ g/ml. The correlation coefficient of the calibration curve was ≥ 0.995 . The concentration of FimCH was determined using regression analysis from a standard curve of the area under the product peak.

15

Construction of Vector pPW19R

This vector was constructed by inserting an intergenic region derived from pPW14R as a Not1-Nco1 fragment into pPW19 creating an additional lacP/O, RBS, and PelB leader. pPW19 represents a modification of the vector pPW16, which was made by cloning the coding sequence for the c-myc epitope into pPW8 as a Not1-Sal1 linker.

pPW8 is a modification of pPW6, which is a derivative of pPW4, the latter being engineered by insertion of a 60 bp synthetic linker containing a Not1 site, a 4D4 epitope sequence, and a termination codon into pPW2 as a BamH1-Sal1 fragment.

pPW2 was made by cloning the full length gene3 (from pSPLIII (JAS)) into pPW1, the latter prepared by cloning the PelB leader sequence (an amino

terminal sequence directing periplasmic insertion of pectin lysate in *E. coli*) into the vector pUC118.

5 Table 3. Primer Sequences used for Vector Construction.

| | <u>Primer*</u> | <u>Sequence</u> | <u>SEQ</u> | <u>ID</u> |
|----|----------------|---|------------|-----------|
| | <u>NO:</u> | | | |
| 10 | GA1F | 5'-CCTGCCATGGCGGGTGTGGCGCTGGGTGCGA CCCGCGTGATTTATCCGGCAGGGC-3' | 56 | |
| | GA2R | 5'-GGCGTCGACAGATTCTATTATTCCATTACGC CCGTC-3' | 57 | |
| 15 | GA13F | 5'-CACACAGGAAACAGCTATGATTGTAATGAAA ACGAG-3' | 58 | |
| | GA6R | 5'-GGCGTCGACGGATCCTTATTGATAAACAAAA GTCACGCC-3' | 59 | |
| 20 | GA11F | 5'-CCGAATAAAGATATCACGACAGGTTTCCCG-3' | 60 | |
| | GA9R | 5'-CATAGCTGTTTCCTGTGTG-3' | 61 | |
| | GA24F | 5'-TGCTCACATGTTCTTTCCTGCGT-3' | 62 | |
| 25 | GA23R | 5'-GACGTTTTTATTACTCATAGCTGTTTCCTGTGTG-3' | 63 | |
| | GA21F | 5'-ATGAGTAATAAAAACGTCAATGTAAGGAAA TCGCAGG-3' | 64 | |

30

* "F" refers to forward primers and "R" refers to reverse primers.

The nucleotide sequences corresponding to the structures used in cloning FimCH were as follows: SEQ ID NO: 47 shows J96 FimC plus a signal sequence and SEQ ID NO: 48 shows J96 FimH; SEQ ID NO: 49 shows the sequence for the kanamycin resistance gene; SEQ ID NO: 50 shows the sequence for *lac I^q*; SEQ ID NO: 51 shows the sequence of Beta-lactamase (with the deletions described in Example 2 covering residues 285 - 335), SEQ ID NO: 52 shows the sequence of the origin of replication, and SEQ ID NO: 53 shows the sequence of Lac p/o (where the two promoter sequences are found at (1) residues 100 to 215, and (2) residues 990 to 1105 of the plasmid (SEQ ID NO: 46)).

Thus, the present invention also relates to a new method of preparing polypeptides from recombinant cells using a vector comprising the plasmid of Figures 3 through 6. In specific embodiments, the polypeptides comprise the amino acid sequences of Figure 2. In one embodiment, this process and plasmid can be used to prepare polypeptides that comprise a bacterial chaperone, such as FimC, fused to a bacterial adhesin, such as FimH or any of the polypeptides presented in Figure 2, including any consensus sequence derived therefrom. The plasmid based procedure of the invention can be used to prepare any polypeptides but especially finds use in preparing the chaperone complexes of the adhesin proteins presented herein. Thus, for example, large amounts of FimCH, the complex of the periplasmic bacterial chaperone, FimC, and the bacterial adhesin, FimH, are readily prepared using the processes disclosed herein.

Numerous modifications and variants of the invention are possible in light of the above teachings; therefore, within the scope of the appended claims, the invention may be practiced otherwise than as specifically described.

WHAT IS CLAIMED IS:

1. An immunogenic polypeptide comprising a sequence selected from the group consisting of residues 26 to 186 of SEQ ID NO: 23 to 45, and 55.
5
2. The immunogenic polypeptide of claim 1 further comprising about the N-terminal two thirds of the sequences selected from SEQ ID NO: 23 to 45, and 55.
- 10 3. The immunogenic polypeptide of claim 1 wherein said polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO: 23 to 45, and 55.
- 15 4. The immunogenic polypeptide of claim 1 wherein said sequence is SEQ ID NO: 55.
5. The immunogenic polypeptide of claim 2 wherein said sequence is SEQ ID NO: 55.
- 20 6. The immunogenic polypeptide of claim 3 wherein said sequence is SEQ ID NO: 55.
7. An isolated polynucleotide encoding a polypeptide of claim 1.
- 25 8. An isolated polynucleotide encoding a polypeptide of claim 3.
9. An isolated polynucleotide encoding a polypeptide of claim 6.
- 30 10. A vaccine composition comprising an immunogenically effective amount of an immunogenic polypeptide of claims 1, 2, or 3, wherein said immunogenic polypeptide is in a pharmacologically acceptable carrier.

11. A vaccine composition comprising an immunogenically effective amount of an immunogenic polypeptide of claims 4, 5, or 6, wherein said immunogenic polypeptide is in a pharmacologically acceptable carrier.

5 12. A vaccine composition comprising an immunogenically effective amount of an immunogenic polypeptide of claims 6, wherein said immunogenic polypeptide is in a pharmacologically acceptable carrier.

10 13. An antibody that binds to a polypeptide having an amino acid sequence selected from SEQ ID NO: 23 to 45, and 55.

14. The antibody of claim 13 wherein said antibody is a monoclonal antibody.

15 15. A process for protecting against an enterobacillus-related disease in a patient at risk of contracting such disease comprising administering to said patient an effective amount of the vaccine composition of claim 10, 11 and 12.

20 16. The method of claim 15 wherein said disease is a urinary tract infection.

17. The method of claim 15 wherein the disease is a bladder infection.

25 18. The method of claim 15 wherein the disease is caused by a bacterium of the family Enterobacteriaceae.

19. The method of claim 18 wherein the bacterium is *E. coli*.

30 20. A process for treating an enterobacillus-related disease in a patient afflicted therewith comprising administering to said patient animal an effective amount of a vaccine composition of claim 10, 11 or 12.

21. The method of claim 20 wherein said disease is a urinary tract infection.

22. The method of claim 20 wherein the disease is a bladder infection.

5

23. The method of claim 20 wherein the disease is caused by a bacterium of the family Enterobacteriaceae.

24. The method of claim 23 wherein the bacterium is *E. coli*.

10

25. A process for treating an enterobacillus-related disease in a patient afflicted therewith comprising administering to said patient an effective amount of an antibody of claim 13 or 14.

15

26. The method of claim 25 wherein said disease is a urinary tract infection.

27. The method of claim 25 wherein the disease is a bladder infection.

20

28. The method of claim 25 wherein the disease is caused by a bacterium of the family Enterobacteriaceae.

29. The method of claim 28 wherein the bacterium is *E. coli*.

25

30. A recombinant cell expressing a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 23 to 45, and 55.

31. The recombinant cell of claim 30 wherein the amino acid sequence is the sequence of SEQ ID NO: 55.

30

32. A vector comprising a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 23 to 45, and 55.

5 33. The vector of claim 32 wherein the amino acid sequence is the sequence of SEQ ID NO: 55.

34. The vector of claim 32 wherein said vector comprises the plasmid pCGA139-1-1.

10

35. The vector of claim 33 wherein said vector comprises the plasmid pCGA139-1-1.

15 36. A process for producing a polypeptide comprising expressing said polypeptide from a recombinant cell containing the vector of claim 34.

37. The process of claim 36 wherein said polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 23 to 45, and 55.

20

38. The process of claim 37 wherein said polypeptide further comprises a bacterial chaperone fused to a bacterial adhesin.

Figure 1(a)

| | |
|-------|---|
| | 90 |
| | 1 |
| J96 | (1) TTGCGCTGTAAAA CCGCCAATGGTACCGCTATCCCTATTGGCGGTGGCAGCGCCCAATGTTTATGTAAACCTTGGCGCCCGTCGTGAATGTG |
| EC45 | (1) TTGCGCTGTAAAA CCGCCAATGGTACCGCAATCCCTATTGGCGGTGGCAGCGCCCAATGTTTATGTAAACCTTGGCGCCCGTCGTGAATGTG |
| B217 | (1) TTGCGCTGTAAAA CCGCCAATGGTACAGCTATCCCTATTGGCGGTGGCAGCGCTAATGTTTATGTAAACCTTGGCGCCCGTCGTGAATGTG |
| DS17 | (1) TTGCGCTGTAAAA CCGCCAATGGTACCGCAATCCCTATTGGCGGTGGCAGCGCCCAATGTTTATGTAAACCTTGGCGCCCGTCGTGAATGTG |
| B212 | (1) TTGCGCTGTAAAA CCGCCAATGGTACCGCTATCCCTATTGGCGGTGGCAGCGCCCAATGTTTATGTAAACCTTGGCGCCCGTCGTGAATGTG |
| EC42 | (1) TTGCGCTGTAAAA CCGCCAATGGTACCGCTATCCCTATTGGCGGTGGCAGCGCCCAATGTTTATGTAAACCTTGGCGCCCGTCGTGAATGTG |
| EC56 | (1) TTGCGCTGTAAAA CCGCCAATGGTACCGCTATCCCTATTGGCGGTGGCAGCGCCCAATGTTTATGTAAACCTTGGCGCCCGTCGTGAATGTG |
| B210 | (1) TTGCGCTGTAAAA CCGCCAATGGTACCGCTATCCCTATTGGCGGTGGCAGCGCCCAATGTTTATGTAAACCTTGGCGCCCGTCGTGAATGTG |
| B203 | (1) TTGCGCTGTAAAA CCGCCAATGGTACCGCAATCCCTATTGGCGGTGGCAGCGCCCAATGTTTATGTAAACCTTGGCGCCCGTCGTGAATGTG |
| EC58 | (1) TTGCGCTGTAAAA CCGCCAATGGTACCGCAATCCCTATTGGCGGTGGCAGCGCCCAATGTTTATGTAAACCTTGGCGCCCGTCGTGAATGTG |
| EC60 | (1) TTGCGCTGTAAAA CCGCCAATGGTACCGCTATCCCTATTGGCGGTGGCAGCGCCCAATGTTTATGTAAACCTTGGCGCCCGTCGTGAATGTG |
| EC61 | (1) TTGCGCTGTAAAA CCGCCAATGGTACCGCTATCCCTATTGGCGGTGGCAGCGCCCAATGTTTATGTAAACCTTGGCGCCCGTCGTGAATGTG |
| EC80 | (1) TTGCGCTGTAAAA CCGCCAATGGTACCGCTATCCCTATTGGCGGTGGCAGCGCCCAATGTTTATGTAAACCTTGGCGCCCGTCGTGAATGTG |
| EC95 | (1) TTGCGCTGTAAAA CCGCCAATGGTACCGCAATCCCTATTGGCGGTGGCAGCGCCCAATGTTTATGTAAACCTTGGCGCCCGTCGTGAATGTG |
| EC62 | (1) ATCGCTGTAAAA CCGCCAATGGTACCGCTATCCCTATTGGCGGTGGCAGCGCCCAATGTTTATGTAAACCTTGGCGCCCGTCGTGAATGTG |
| B238 | (1) TTGCGCTGTAAAA CCGCCAATGGTACCGCTATCCCTATTGGCGGTGGCAGCGCCCAATGTTTATGTAAACCTTGGCGCCCGTCGTGAATGTG |
| B240 | (1) TTGCGCTGTAAAA CCGCCAATGGTACCGCTATCCCTATTGGCGGTGGCAGCGCCCAATGTTTATGTAAACCTTGGCGCCCGTCGTGAATGTG |
| B242 | (1) TTGCGCTGTAAAA CCGCCAATGGTACCGCTATCCCTATTGGCGGTGGCAGCGCCCAATGTTTATGTAAACCTTGGCGCCCGTCGTGAATGTG |
| EC189 | (1) TTGCGCTGTAAAA CCGCCAATGGTACCGCTATCCCTATTGGCGGTGGCAGCGCCCAATGTTTATGTAAACCTTGGCGCCCGTCGTGAATGTG |

Figure 1(b)

| | | | | | |
|-------|------|----------|--|--|-----|
| J96 | (91) | GGGCAAAA | CCTGGTCGTGGATCTTTTCGACGCAAAATCTTTTGGCCATAA | CGAATTATCCGGAAACCAATTACAGACTATGTCACTGCAACCGA | 180 |
| EC45 | (91) | GGGCAAAA | CCTGGTCGTAGATCTTTTCGACGCAAAATCTTTTGGCCATAA | CGAATTATCCAGAAACCAATTACAGACTATGTCACTGCAACCGA | |
| B217 | (91) | GGGCAAAA | CCTGGTCGTAGATCTTTTCGACGCAAAATCTTTTGGCCATAA | CGAATTATCCGGAAACCAATTACAGACTATGTCACTGCAACCGA | |
| DS17 | (91) | GGGCAAAA | CCTGGTCGTAGATCTTTTCGACGCAAAATCTTTTGGCCATAA | CGAATTATCCAGAAACCAATTACAGACTATGTCACTGCAACCGA | |
| B212 | (91) | GGGCAAAA | CCTGGTCGTGGATCTTTTCGACGCAAAATCTTTTGGCCATAA | CGAATTATCCGGAAACCAATTACAGACTATGTCACTGCAACCGA | |
| EC42 | (91) | GGGCAAAA | CCTGGTCGTGGATCTTTTCGACGCAAAATCTTTTGGCCATAA | CGAATTATCCGGAAACCAATTACAGACTATGTCACTGCAACCGA | |
| EC56 | (91) | GGGCAAAA | CCTGGTCGTGGATCTTTTCGACGCAAAATCTTTTGGCCATAA | CGAATTATCCGGAAACCAATTACAGACTATGTCACTGCAACCGA | |
| B210 | (91) | GGGCAAAA | CCTGGTCGTGGATCTTTTCGACGCAAAATCTTTTGGCCATAA | CGAATTATCCGGAAACCAATTACAGACTATGTCACTGCAACCGA | |
| B203 | (91) | GGGCAAAA | CCTGGTCGTGGATCTTTTCGACGCAAAATCTTTTGGCCATAA | CGAATTATCCGGAAACCAATTACAGACTATGTCACTGCAACCGA | |
| EC58 | (91) | GGGCAAAA | CCTGGTCGTGGATCTTTTCGACGCAAAATCTTTTGGCCATAA | CGAATTATCCGGAAACCAATTACAGACTATGTCACTGCAACCGA | |
| EC60 | (91) | GGGCAAAA | CCTGGTCGTGGATCTTTTCGACGCAAAATCTTTTGGCCATAA | CGAATTATCCGGAAACCAATTACAGACTATGTCACTGCAACCGA | |
| EC61 | (91) | GGGCAAAA | CCTGGTCGTGGATCTTTTCGACGCAAAATCTTTTGGCCATAA | CGAATTATCCGGAAACCAATTACAGACTATGTCACTGCAACCGA | |
| EC80 | (91) | GGGCAAAA | CCTGGTCGTGGATCTTTTCGACGCAAAATCTTTTGGCCATAA | CGAATTATCCGGAAACCAATTACAGACTATGTCACTGCAACCGA | |
| EC95 | (91) | GGGCAAAA | CCTGGTCGTGGATCTTTTCGACGCAAAATCTTTTGGCCATAA | CGAATTATCCGGAAACCAATTACAGACTATGTCACTGCAACCGA | |
| EC62 | (91) | GGGCAAAA | CCTGGTCGTGGATCTTTTCGACGCAAAATCTTTTGGCCATAA | CGAATTATCCGGAAACCAATTACAGACTATGTCACTGCAACCGA | |
| B238 | (91) | GGGCAAAA | CCTGGTCGTGGATCTTTTCGACGCAAAATCTTTTGGCCATAA | CGAATTATCCGGAAACCAATTACAGACTATGTCACTGCAACCGA | |
| B240 | (91) | GGGCAAAA | CCTGGTCGTGGATCTTTTCGACGCAAAATCTTTTGGCCATAA | CGAATTATCCGGAAACCAATTACAGACTATGTCACTGCAACCGA | |
| B242 | (91) | GGGCAAAA | CCTGGTCGTGGATCTTTTCGACGCAAAATCTTTTGGCCATAA | CGAATTATCCGGAAACCAATTACAGACTATGTCACTGCAACCGA | |
| EC189 | (91) | GGGCAAAA | CCTGGTCGTGGATCTTTTCGACGCAAAATCTTTTGGCCATAA | CGAATTATCCGGAAACCAATTACAGACTATGTCACTGCAACCGA | |

Figure 1(c)

181
 J96 (181) GGCTCGGCTTATGGCGGCGTGTATCTAAATTTTCCGGGACCGTAAATATATAGTGGCAGTAGCTATCCATTTCCCTACCAACGCGGAAACG
 EC45 (181) GGTCGGCTTATGGCGGCGTGTATCTAGTTTTCGGGACCGTAAATATATAGTGGCAGTAGCTATCCATTTCCCTACTACGAGCGAAACG
 B217 (181) GGCTCGGCTTATGGCGGCGTGTATCTAAATTTTCCGGGACCGTAAATATATAGTGGCAGTAGCTATCCATTTCCGACCAACGCGGAAACG
 DS17 (181) GGTCGGCTTATGGCGGCGTGTATCTAGTTTTCGGGACCGTAAATATATAGTGGCAGTAGCTATCCATTTCCCTACTACGAGCGAAACG
 B212 (181) GGTCGGCTTATGGCGGCGTGTATCTAGTTTTCGGGACCGTAAATATATAGTGGCAGTAGCTATCCATTTCCGACCAACGAGTGAACG
 EC42 (181) GGCTCGGCTTATGGCGGCGTGTATCTAAATTTTCCGGGACCGTAAATATATAGTGGCAGTAGCTATCCATTTCCCTACCAACGCGGAAACG
 EC56 (181) GGCTCGGCTTATGGCGGCGTGTATCTAAATTTTCCGGGACCGTAAATATATAGTGGCAGTAGCTATCCATTTCCCTACCAACGCGGAAACG
 B210 (181) GGCTCGGCTTATGGCGGCGTGTATCTAAATTTTCCGGGACCGTAAATATATAGTGGCAGTAGCTATCCATTTCCCTACCAACGCGGAAACG
 B203 (181) GGTCGGCTTATGGCGGCGTGTATCTAGTTTTCGGGACCGTAAATATATAGTGGCAGTAGCTATCCATTTCCCTACTACGAGCGAAACG
 EC58 (181) GGTCGGCTTATGGCAGCGTGTATCTAGTTTTCGGGACCGTAAATATATAGTGGCAGTAGCTATCCATTTCCGACCAACGAGTGAACG
 EC60 (181) GGCTCGGCTTATGGCGGCGTGTATCTAAATTTTCCGGGACCGTAAATATATAGTGGCAGTAGCTATCCATTTCCGACCAACGAGTGAACG
 EC61 (181) GGCTCGGCTTATGGCGGCGTGTATCTAAATTTTCCGGGACCGTAAATATATAGTGGCAGTAGCTATCCATTTCCCTACTACGAGCGAAACG
 EC80 (181) GGCTCGGCTTATGGCGGCGTGTATCTAAATTTTCCGGGACCGTAAATATATAGTGGCAGTAGCTATCCATTTCCCTACTACGAGCGAAACG
 EC95 (181) GGTCGGCTTATGGCGGCGTGTATCTAGTTTTCGGGACCGTAAATATATAGTGGCAGTAGCTATCCATTTCCCTACTACGAGCGAAACG
 EC62 (181) GGTCGGCTTATGGCGGCGTGTATCTCAATTTTCCGGGACCGTAAATATATAGTGGCAGTAGCTATCCATTTCCCTACTACGAGCGAAACG
 B238 (181) GGCTCGGCTTATGGCGGCGTGTATCTAAATTTTCCGGGACCGTAAATATATAGTGGCAGTAGCTATCCATTTCCGACCAACGAGTGAACG
 B240 (181) GGCTCGGCTTATGGCGGCGTGTATCTAAATTTTCCGGGACCGTAAATATATAGTGGCAGTAGCTATCCATTTCCCTACTACGAGCGAAACG
 B242 (181) GGCTCGGCTTATGGCGGCGTGTATCTAAATTTTCCGGGACCGTAAATATATAGTGGCAGTAGCTATCCATTTCCGACTACGAGCGAAACG
 EC189 (181) GGCTCGGCTTATGGCGGCGTGTATCTAAATTTTCCGGGACCGTAAATATATAGTGGCAGTAGCTATCCATTTCCGACCAACGAGTGAACG

Figure 1(d)

271 (271) CCGCGGTTGTTTATAATTTCGAGAACGGATAAGCCGTGGCCGGTGGCGCTTTATTGACGCCCTGTGAGCAGTGC CGGCGGGGTGGCGATT
 J96 (271) CCGCGGTTGTTTATAATTTCGAGAACGGATAAGCCGTGGCCGGTGGCGCTTTATTGACGCCCTGTGAGCAGTGC CGGCGGGGTGGCGATT
 EC45 (271) CCGCGGTTGTTTATAATTTCGAGAACGGATAAGCCGTGGCCGGTGGCGCTTTATTGACGCCCTGTGAGCAGTGC CGGCGGGGTGGCGATT
 B217 (271) CCGCGGTTGTTTATAATTTCGAGAACGGATAAGCCGTGGCCGGTGGCGCTTTATTGACGCCCTGTGAGCAGTGC CGGCGGGGTGGCGATT
 DS17 (271) CCGCGGTTGTTTATAATTTCGAGAACGGATAAGCCGTGGCCGGTGGCGCTTTATTGACGCCCTGTGAGCAGTGC CGGCGGGGTGGCGATT
 B212 (271) CCGCGGTTGTTTATAATTTCGAGAACGGATAAGCCGTGGCCGGTGGCGCTTTATTGACGCCCTGTGAGCAGTGC CGGCGGGGTGGCGATT
 EC42 (271) CCGCGGTTGTTTATAATTTCGAGAACGGATAAGCCGTGGCCGGTGGCGCTTTATTGACGCCCTGTGAGCAGTGC CGGCGGGGTGGCGATT
 EC56 (271) CCGCGGTTGTTTATAATTTCGAGAACGGATAAGCCGTGGCCGGTGGCGCTTTATTGACGCCCTGTGAGCAGTGC CGGCGGGGTGGCGATT
 B210 (271) CCGCGGTTGTTTATAATTTCGAGAACGGATAAGCCGTGGCCGGTGGCGCTTTATTGACGCCCTGTGAGCAGTGC CGGCGGGGTGGCGATT
 B203 (271) CCGCGGTTGTTTATAATTTCGAGAACGGATAAGCCGTGGCCGGTGGCGCTTTATTGACGCCCTGTGAGCAGTGC CGGCGGGGTGGCGATT
 EC58 (271) CCGCGGTTGTTTATAATTTCGAGAACGGATAAGCCGTGGCCGGTGGCGCTTTATTGACGCCCTGTGAGCAGTGC CGGCGGGGTGGCGATT
 EC60 (271) CCGCGGTTGTTTATAATTTCGAGAACGGATAAGCCGTGGCCGGTGGCGCTTTATTGACGCCCTGTGAGCAGTGC CGGCGGGGTGGCGATT
 EC61 (271) CCGCGGTTGTTTATAATTTCGAGAACGGATAAGCCGTGGCCGGTGGCGCTTTATTGACGCCCTGTGAGCAGTGC CGGCGGGGTGGCGATT
 EC80 (271) CCGCGGTTGTTTATAATTTCGAGAACGGATAAGCCGTGGCCGGTGGCGCTTTATTGACGCCCTGTGAGCAGTGC CGGCGGGGTGGCGATT
 EC95 (271) CCGCGGTTGTTTATAATTTCGAGAACGGATAAGCCGTGGCCGGTGGCGCTTTATTGACGCCCTGTGAGCAGTGC CGGCGGGGTGGCGATT
 EC62 (271) CCGCGGTTGTTTATAATTTCGAGAACGGATAAGCCGTGGCCGGTGGCGCTTTATTGACGCCCTGTGAGCAGTGC CGGCGGGGTGGCGATT
 B238 (271) CCGCGGTTGTTTATAATTTCGAGAACGGATAAGCCGTGGCCGGTGGCGCTTTATTGACGCCCTGTGAGCAGTGC CGGCGGGGTGGCGATT
 B240 (271) CCGCGGTTGTTTATAATTTCGAGAACGGATAAGCCGTGGCCGGTGGCGCTTTATTGACGCCCTGTGAGCAGTGC CGGCGGGGTGGCGATT
 B242 (271) CCGCGGTTGTTTATAATTTCGAGAACGGATAAGCCGTGGCCGGTGGCGCTTTATTGACGCCCTGTGAGCAGTGC CGGCGGGGTGGCGATT
 EC189 (271) CCGCGGTTGTTTATAATTTCGAGAACGGATAAGCCGTGGCCGGTGGCGCTTTATTGACGCCCTGTGAGCAGTGC CGGCGGGGTGGCGATT

Figure 1(e)

361 450

J96 (361) AAAGCTGGGCTCAATTAATTGCCGTGCTTAATTTTGGGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTGTGGAATATTTACGCC
EC45 (361) AAAGCTGGGCTCAATTAATTGCCGTGCTTAATTTTGGGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTGTGGAATATTTACGCC
B217 (361) AAAGCTGGGCTCAATTAATTGCCGTGCTTAATTTTGGGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTGTGGAATATTTACGCC
DS17 (361) AAAGCTGGGCTCAATTAATTGCCGTGCTTAATTTTGGGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTGTGGAATATTTACGCC
B212 (361) AAAGCAGGCTCAATTAATTGCCGTGCTTAATTTTGGGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTGTGGAATATTTACGCC
EC42 (361) AAAGCTGGGCTCAATTAATTGCCGTGCTTAATTTTGGGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTGTGGAATATTTACGCC
EC56 (361) AAAGCTGGGCTCAATTAATTGCCGTGCTTAATTTTGGGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTGTGGAATATTTACGCC
B210 (361) AAAGCAGGCTCAATTAATTGCCGTGCTTAATTTTGGGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTGTGGAATATTTACGCC
B203 (361) AAAGCTGGGCTCAATTAATTGCCGTGCTTAATTTTGGGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTGTGGAATATTTACGCC
EC58 (361) AAAGCTGGGCTCAATTAATTGCCGTGCTTAATTTTGGGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTGTGGAATATTTACGCC
EC60 (361) AAAGCTGGGCTCAATTAATTGCCGTGCTTAATTTTGGGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTGTGGAATATTTACGCC
EC61 (361) AAAGCTGGGCTCAATTAATTGCCGTGCTTAATTTTGGGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTGTGGAATATTTACGCC
EC80 (361) AAAGCTGGGCTCAATTAATTGCCGTGCTTAATTTTGGGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTGTGGAATATTTACGCC
EC95 (361) AAAGCTGGGCTCAATTAATTGCCGTGCTTAATTTTGGGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTGTGGAATATTTACGCC
EC62 (361) AAAGCTGGGCTCAATTAATTGCCGTGCTTAATTTTGGGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTGTGGAATATTTACGCC
B238 (361) AAAGCTGGGCTCAATTAATTGCCGTGCTTAATTTTGGGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTGTGGAATATTTACGCC
B240 (361) AAAGCTGGGCTCAATTAATTGCCGTGCTTAATTTTGGGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTGTGGAATATTTACGCC
B242 (361) AAAGCTGGGCTCAATTAATTGCCGTGCTTAATTTTGGGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTGTGGAATATTTACGCC
EC189 (361) AAAGCTGGGCTCAATTAATTGCCGTGCTTAATTTTGGGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTGTGGAATATTTACGCC

Figure 1(f)

| | | |
|-------|-------|---|
| | 451 | 540 |
| J96 | (451) | AATAATGATGTGGTGGTGCCTACTGCGCGCTGCGATGTTTCTGCTCGTGATGTACCGTTACTCTGCGCGGACTACCCCTGGTTCAGTGCCA |
| EC45 | (451) | AATAATGATGTGGTGGTGCCTACTGCGCGCTGCGATGTTTCTGCTCGTGATGTACCGTTACTCTGCGCGGACTACCCCTGGTTCAGTGCCG |
| B217 | (451) | AATAATGATGTGGTGGTGCCTACTGCGCGCTGCGATGTTTCTGCTCGTGATGTACCGTTACTCTGCGCGGACTACCCCTGGTTCAGTGCCA |
| DS17 | (451) | AATAATGATGTGGTGGTGCCTACTGCGCGCTGCGATGTTTCTGCTCGTGATGTACCGTTACTCTGCGCGGACTACCCCTGGTTCAGTGCCG |
| B212 | (451) | AATAATGATGTGGTGGTGCCTACTGCGCGCTGCGATGTTTCTGCTCGTGATGTACCGTTACTCTGCGCGGACTACCCCTGGTTCAGTGCCG |
| EC42 | (451) | AATAATGATGTGGTGGTGCCTACTGCGCGCTGCGATGTTTCTGCTCGTGATGTACCGTTACTCTGCGCGGACTACCCCTGGTTCAGTGCCG |
| EC56 | (451) | AATAATGATGTGGTGGTGCCTACTGCGCGCTGCGATGTTTCTGCTCGTGATGTACCGTTACTCTGCGCGGACTACCCCTGGTTCAGTGCCA |
| B210 | (451) | AATAATGATGTGGTGGTGCCTACTGCGCGCTGCGATGTTTCTGCTCGTGATGTACCGTTACTCTGCGCGGACTACCCCTGGTTCAGTGCCG |
| B203 | (451) | AATAATGATGTGGTGGTGCCTACTGCGCGCTGCGATGTTTCTGCTCGTGATGTACCGTTACTCTGCGCGGACTACCCCTGGTTCAGTGCCG |
| EC58 | (451) | AATAATGATGTGGTGGTGCCTACTGCGCGCTGCGATGTTTCTGCTCGTGATGTACCGTTACTCTGCGCGGACTACCCCTGGTTCAGTGCCG |
| EC60 | (451) | AATAATGATGTGGTGGTGCCTACTGCGCGCTGCGATGTTTCTGCTCGTGATGTACCGTTACTCTGCGCGGACTACCCCTGGTTCAGTGCCG |
| EC61 | (451) | AATAATGATGTGGTGGTGCCTACTGCGCGCTGCGATGTTTCTGCTCGTGATGTACCGTTACTCTGCGCGGACTACCCCTGGTTCAGTGCCA |
| EC80 | (451) | AATAATGATGTGGTGGTGCCTACTGCGCGCTGCGATGTTTCTGCTCGTGATGTACCGTTACTCTGCGCGGACTACCCCTGGTTCAGTGCCA |
| EC95 | (451) | AATAATGATGTGGTGGTGCCTACTGCGCGCTGCGATGTTTCTGCTCGTGATGTACCGTTACTCTGCGCGGACTACCCCTGGTTCAGTGCCG |
| EC62 | (451) | AATAATGATGTGGTGGTGCCTACTGCGCGCTGCGATGTTTCTGCTCGTGATGTACCGTTACTCTGCGCGGACTACCCCTGGTTCAGTGCCG |
| B238 | (451) | AATAATGATGTGGTGGTGCCTACTGCGCGCTGCGATGTTTCTGCTCGTGATGTACCGTTACTCTGCGCGGACTACCCCTGGTTCAGTGCCG |
| B240 | (451) | AATAATGATGTGGTGGTGCCTACTGCGCGCTGCGATGTTTCTGCTCGTGATGTACCGTTACTCTGCGCGGACTACCCCTGGTTCAGTGCCA |
| B242 | (451) | AATAATGATGTGGTGGTGCCTACTGCGCGCTGCGATGTTTCTGCTCGTGATGTACCGTTACTCTGCGCGGACTACCCCTGGTTCAGTGCCA |
| EC189 | (451) | AATAATGATGTGGTGGTGCCTACTGCGCGCTGCGATGTTTCTGCTCGTGATGTACCGTTACTCTGCGCGGACTACCCCTGGTTCAGTGCCA |

Figure 1 (g)

541 (541) ATTCTCTTACCGTTTATTGTGCGAAAGCCAAACCTGGGTATTACCTCTCCGGCAAAACCGCAGATGCGGGCAACTCGATTTTCACC
 J96 (541) ATTCTCTTACCGTTTATTGTGCGAAAGCCAAACCTGGGTATTACCTCTCCGGCAAAACCGCAGATGCGGGCAACTCGATTTTCACC
 EC45 (541) ATTCTCTTACCGTTTATTGTGCGAAAGCCAAACCTGGGTATTACCTCTCCGGCAAAACCGCAGATGCGGGCAACTCGATTTTCACC
 B217 (541) ATTCTCTTACCGTTTATTGTGCGAAAGCCAAACCTGGGTATTACCTCTCCGGCAAAACCGCAGATGCGGGCAACTCGATTTTCACC
 DS17 (541) ATTCTCTTACCGTTTATTGTGCGAAAGCCAAACCTGGGTATTACCTCTCCGGCAAAACCGCAGATGCGGGCAACTCGATTTTCACC
 B212 (541) ATTCTCTTACCGTTTATTGTGCGAAAGCCAAACCTGGGTATTACCTCTCCGGCAAAACCGCAGATGCGGGCAACTCGATTTTCACC
 EC42 (541) ATTCTCTTACCGTTTATTGTGCGAAAGCCAAACCTGGGTATTACCTCTCCGGCAAAACCGCAGATGCGGGCAACTCGATTTTCACC
 EC56 (541) ATTCTCTTACCGTTTATTGTGCGAAAGCCAAACCTGGGTATTACCTCTCCGGCAAAACCGCAGATGCGGGCAACTCGATTTTCACC
 B210 (541) ATTCTCTTACCGTTTATTGTGCGAAAGCCAAACCTGGGTATTACCTCTCCGGCAAAACCGCAGATGCGGGCAACTCGATTTTCACC
 B203 (541) ATTCTCTTACCGTTTATTGTGCGAAAGCCAAACCTGGGTATTACCTCTCCGGCAAAACCGCAGATGCGGGCAACTCGATTTTCACC
 EC58 (541) ATTCTCTTACCGTTTATTGTGCGAAAGCCAAACCTGGGTATTACCTCTCCGGCAAAACCGCAGATGCGGGCAACTCGATTTTCACC
 EC60 (541) ATTCTCTTACCGTTTATTGTGCGAAAGCCAAACCTGGGTATTACCTCTCCGGCAAAACCGCAGATGCGGGCAACTCGATTTTCACC
 EC61 (541) ATTCTCTTACCGTTTATTGTGCGAAAGCCAAACCTGGGTATTACCTCTCCGGCAAAACCGCAGATGCGGGCAACTCGATTTTCACC
 EC80 (541) ATTCTCTTACCGTTTATTGTGCGAAAGCCAAACCTGGGTATTACCTCTCCGGCAAAACCGCAGATGCGGGCAACTCGATTTTCACC
 EC95 (541) ATTCTCTTACCGTTTATTGTGCGAAAGCCAAACCTGGGTATTACCTCTCCGGCAAAACCGCAGATGCGGGCAACTCGATTTTCACC
 EC62 (541) ATTCTCTTACCGTTTATTGTGCGAAAGCCAAACCTGGGTATTACCTCTCCGGCAAAACCGCAGATGCGGGCAACTCGATTTTCACC
 B238 (541) ATTCTCTTACCGTTTATTGTGCGAAAGCCAAACCTGGGTATTACCTCTCCGGCAAAACCGCAGATGCGGGCAACTCGATTTTCACC
 B240 (541) ATTCTCTTACCGTTTATTGTGCGAAAGCCAAACCTGGGTATTACCTCTCCGGCAAAACCGCAGATGCGGGCAACTCGATTTTCACC
 B242 (541) ATTCTCTTACCGTTTATTGTGCGAAAGCCAAACCTGGGTATTACCTCTCCGGCAAAACCGCAGATGCGGGCAACTCGATTTTCACC
 EC189 (541) ATTCTCTTACCGTTTATTGTGCGAAAGCCAAACCTGGGTATTACCTCTCCGGCAAAACCGCAGATGCGGGCAACTCGATTTTCACC

Figure 1 (h)

| | | | | |
|-------|-------|--------------------------|---------------------------|---|
| J96 | (631) | AATACCGCGTCGTTTTCACCTGCA | CAGGGCGTCGGCGTACAGTTGAC | CGCGCAACGGTACGATTATTTCCAGCGAATAACACGGTATCGTTA |
| EC45 | (631) | AATACCGCGTCGTTTTCACCCGCG | CAGGGCGTCGGCGTACAGTTGAC | GGCGCAACGGTACGATTATTTCCAGCGAATAACACGGTATCGTTA |
| B217 | (631) | AATACCGCGTCGTTTTCACAGCG | CAGGGCGTCGGCGTTCAGTTGAC | GGCGCAACGGTACGATTATTTCCCA |
| DS17 | (631) | AATACCGCGTCGTTTTCACCCGCG | CAGGGCGTCGGCGTACAGTTGAC | GGCGCAACGGTACGATTATTTCCAGCGAATAACACGGTATCGTTA |
| B212 | (631) | AATACCGCGTCGTTTTCACCCGCG | CAGGGCGTCGGCGTACAGTTGAC | GGCGCAACGGTACGATTATTTCCAGCGAATAACACGGTATCGTTA |
| EC42 | (631) | AATACCGCGTCGTTTTCACCTGCA | CAGGGCGTCGGCGTACAGTTGAC | GGCGCAACGGTACGATTATTTCCAGCGAATAACACGGTATCGTTA |
| EC56 | (631) | AATACCGCGTCGTTTTCACCTGCA | CAGGGCGTCGGCGTACAGTTGAC | GGCGCAACGGTACGATTATTTCCAGCGAATAACACGGTATCGTTA |
| B210 | (631) | AATACCGCGTCGTTTTCACCCGCG | CAGGGCGTCGGCGTACAGTTGAC | GGCGCAACGGTACGATTATTTCCAGCGAATAACACGGTATCGTTA |
| B203 | (631) | AATACCGCGTCGTTTTCACCCGCG | CAGGGCGTCGGCGTACAGTTGAC | GGCGCAACGGTACGATTATTTCCAGCGAATAACACGGTATCGTTA |
| EC58 | (631) | AATACCGCGTCGTTTTCACCCGCG | CAGGGCGTCGGCGTACAGTTGAC | GGCGCAACGGTACGATTATTTCCAGCGAATAACACGGTATCGTTA |
| EC60 | (631) | AATACCGCGTCGTTTTCACCTGCA | CAGGGCGTCGGCGTACAGTTGAC | GGCGCAACGGTACGATTATTTCCAGCGAATAACACGGTATCGTTA |
| EC61 | (631) | AATACCGCGTCGTTTTCACCTGCA | CAGGGCGTCGGCGTACAGTTGAC | GGCGCAACGGTACGATTATTTCCAGCGAATAACACGGTATCGTTA |
| EC80 | (631) | AATACCGCGTCGTTTTCACCTGCA | CAGGGCGTCGGCGTACAGTTGAC | GGCGCAACGGTACGATTATTTCCAGCGAATAACACGGTATCGTTA |
| EC95 | (631) | AATACCGCGTCGTTTTCACCCGCG | CAGGGCGTCGGCGTACAGTTGAC | GGCGCAACGGTACGATTATTTCCAGCGAATAACACGGTATCGTTA |
| EC62 | (631) | AATACCGCGTCGTTTTCACCTGCA | CAGGGCGTCGGCGTACAGTTGAC | GGCGCAACGGTACGATTATTTCCAGCGAATAACACGGTATCGTTA |
| B238 | (631) | AATACCGCGTCGTTTTCACCTGCA | CAGGGCGTCGGCGTACAGTTGAC | GGCGCAACGGTACGATTATTTCCAGCGAATAACACGGTATCGTTA |
| B240 | (631) | AATACCGCGTCGTTTTCACCTGCA | CAGGGCGTCGGCGTACAGTTGAC | GGCGCAACGGTACGATTATTTCCAGCGAATAACACGGTATCGTTA |
| B242 | (631) | AATACCGCGTCGTTTTCACAGCG | CAGGGCGTCGGCGTACAGTTGAC | GGCGCAACGGTACGATTATTTCCAGCGAATAACACGGTATCGTTA |
| EC189 | (631) | AATACCGCGTCGTTTTCACCCAG | CGCAGGGCGTCGGCGTACAGTTGAC | GGCGCAACGGTACGATTATTTCCAGCGAATAACACGGTATCGTTA |

Figure 1(i)

J96 (721) GGAGCAGTAGGGACTTCGGCGGTAGTCTGGGATTAA CGGCAAATTATGCA CGTACCGGAGGGCAGGTGACTG CAGGGAAATGTGCAATCG
 EC45 (721) GGAGCAGTAGGGACTTCGGCGGTAGTCTGGGATTAA CGGCAAATTATGCA CGTACCGGAGGGCAGGTGACTG CAGGGAAATGTGCAATCG
 B217 (721) GGAGCAGTAGGGACTTCGGCGGTAGTCTGGGATTAA CGGCAAATTATGCA CGTACCGGAGGGCAGGTGACTG CAGGGAAATGTGCAATCG
 DS17 (721) GGAGCAGTAGGGACTTCGGCGGTAGTCTGGGATTAA CGGCAAATTATGCA CGTACCGGAGGGCAGGTGACTG CAGGGAAATGTGCAATCG
 B212 (721) GGAGCAGTAGGGACTTCGGCGGTAGTCTGGGATTAA CGGCAAATTATGCA CGTACCGGAGGGCAGGTGACTG CAGGGAAATGTGCAATCG
 EC42 (721) GGAGCAGTAGGGACTTCGGCGGTAGTCTGGGATTAA CGGCAAATTATGCA CGTACCGGAGGGCAGGTGACTG CAGGGAAATGTGCAATCG
 EC56 (721) GGAGCAGTAGGGACTTCGGCGGTAGTCTGGGATTAA CGGCAAATTATGCA CGTACCGGAGGGCAGGTGACTG CAGGGAAATGTGCAATCG
 B210 (721) GGAGCAGTAGGGACTTCGGCGGTAGTCTGGGATTAA CGGCAAATTATGCA CGTACCGGAGGGCAGGTGACTG CAGGGAAATGTGCAATCG
 B203 (721) GGAGCAGTAGGGACTTCGGCGGTAGTCTGGGATTAA CGGCAAATTATGCA CGTACCGGAGGGCAGGTGACTG CAGGGAAATGTGCAATCG
 EC58 (721) GGAGCAGTAGGGACTTCGGCGGTAGTCTGGGATTAA CGGCAAATTATGCA CGTACCGGAGGGCAGGTGACTG CAGGGAAATGTGCAATCG
 EC60 (721) GGAGCAGTAGGGACTTCGGCGGTAGTCTGGGATTAA CGGCAAATTATGCA CGTACCGGAGGGCAGGTGACTG CAGGGAAATGTGCAATCG
 EC61 (721) GGAGCAGTAGGGACTTCGGCGGTAGTCTGGGATTAA CGGCAAATTATGCA CGTACCGGAGGGCAGGTGACTG CAGGGAAATGTGCAATCG
 EC80 (721) GGAGCAGTAGGGACTTCGGCGGTAGTCTGGGATTAA CGGCAAATTATGCA CGTACCGGAGGGCAGGTGACTG CAGGGAAATGTGCAATCG
 EC95 (721) GGAGCAGTAGGGACTTCGGCGGTAGTCTGGGATTAA CGGCAAATTATGCA CGTACCGGAGGGCAGGTGACTG CAGGGAAATGTGCAATCG
 EC62 (721) GGAGCAGTAGGGACTTCGGCGGTAGTCTGGGATTAA CGGCAAATTATGCA CGTACCGGAGGGCAGGTGACTG CAGGGAAATGTGCAATCG
 B238 (721) GGAGCAGTAGGGACTTCGGCGGTAGTCTGGGATTAA CGGCAAATTATGCA CGTACCGGAGGGCAGGTGACTG CAGGGAAATGTGCAATCG
 B240 (721) GGAGCAGTAGGGACTTCGGCGGTAGTCTGGGATTAA CGGCAAATTATGCA CGTACCGGAGGGCAGGTGACTG CAGGGAAATGTGCAATCG
 B242 (721) GGAGCAGTAGGGACTTCGGCGGTAGTCTGGGATTAA CGGCAAATTATGCA CGTACCGGAGGGCAGGTGACTG CAGGGAAATGTGCAATCG
 EC189 (721) GGAGCAGTAGGGACTTCGGCGGTAGTCTGGGATTAA CGGCAAATTATGCA CGTACCGGAGGGCAGGTGACTG CAGGGAAATGTGCAATCG

Figure 1(j)

| | | | | |
|-------|-------|-----|-------------------------------|-----|
| J96 | (811) | 811 | ATTATTGGCGTGACCTTTTGTATTATCAA | 837 |
| EC45 | (811) | | ATTATTGGCGTGACCTTTTGTATTATCAA | |
| B217 | (811) | | ATTATTGGCGTGACCTTTTGTATTATCAA | |
| DS17 | (811) | | ATTATTGGCGTGACCTTTTGTATTATCAA | |
| B212 | (811) | | ATTATTGGCGTGACCTTTTGTATTATCAA | |
| EC42 | (811) | | ATTATTGGCGTGACCTTTTGTATTATCAA | |
| EC56 | (811) | | ATTATTGGCGTGACCTTTTGTATTATCAA | |
| B210 | (811) | | ATTATTGGCGTGACCTTTTGTATTATCAA | |
| B203 | (811) | | ATTATTGGCGTGACCTTTTGTATTATCAA | |
| EC58 | (811) | | ATTATTGGCGTGACCTTTTGTATTATCAA | |
| EC60 | (811) | | ATTATTGGCGTGACCTTTTGTATTATCAA | |
| EC61 | (811) | | ATTATTGGCGTGACCTTTTGTATTATCAA | |
| EC80 | (811) | | ATTATTGGCGTGACCTTTTGTATTATCAA | |
| EC95 | (811) | | ATTATTGGCGTGACCTTTTGTATTATCAA | |
| EC62 | (811) | | ATTATTGGCGTGACCTTTTGTATTATCAA | |
| B238 | (811) | | ATTATTGGCGTGACCTTTTGTATTATCAA | |
| B240 | (811) | | ATTATTGGCGTGACCTTTTGTATTATCAA | |
| B242 | (811) | | ATTATTGGCGTGACCTTTTGTATTATCAA | |
| EC189 | (811) | | ATTATTGGCGTGACCTTTTGTATTATCAA | |

Figure 2(a)

| | 1 | 50 |
|-----------|-----|--|
| B210.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPVNVGQNLVVDLSTQIFCHNDYPE |
| B212.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPE |
| B217.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPE |
| B223.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPE |
| B228.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPE |
| B238.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAIAVNVGQNLVVDLSTQIFCHNDYPE |
| B240.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPVNVGQNLVVDLSTQIFCHNDYPE |
| B242.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPE |
| DS17.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPE |
| EC42.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPE |
| EC45.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPE |
| EC56.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPVNVGQNLVVDLSTQIFCHNDYPE |
| EC58.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPE |
| EC60.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPVNVGQNLVVDLSTQIFCHNDYPE |
| EC61.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPE |
| EC62.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPE |
| EC80.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPE |
| EC89.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPE |
| EC95.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPE |
| G189.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPE |
| J96.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPVNVGQNLVVDLSTQIFCHNDYPE |
| NU14.aa | (1) | PACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPE |
| Consensus | (1) | PACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPE |

| | 51 | 100 |
|-----------|------|---|
| B210.aa | (51) | TITDYVTLQRGSAYGGVLSNFSGIVKYSGSSYPFPTTSETPRVVYNSRTD |
| B212.aa | (51) | TITDYVTLQRGSAYGGVLSFSFGIVKYNGSSYPFPTTSETPRVVYNSRTD |
| B217.aa | (51) | TITDYVTLQRGAAYGGVLSFSFGTVKYNGSSYPFPTTSETPRVVYNSRTD |
| B223.aa | (51) | TITDYVTLQRGAAYGGVLSFSFGTVKYNGSSYPFPTTSETPRVVYNSRTD |
| B228.aa | (51) | TITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTD |
| B238.aa | (51) | TITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTD |
| B240.aa | (51) | TITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTD |
| B242.aa | (51) | TITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTD |
| DS17.aa | (51) | TITDYVTLQRGSAYGGVLSFSFGTVKYNGSSYPFPTTSETPRVVYNSRTD |
| EC42.aa | (51) | TITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTD |
| EC45.aa | (51) | TITDYVTLQRGAAYGGVLSFSFGTVKYNGSSYPFPTTSETPRVVYNSRTD |
| EC56.aa | (51) | TITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTD |
| EC58.aa | (51) | TITDYVTLQRGSAYGVLSSFSFGTVKYNGSSYPFPTTSETPRVVYNSRTD |
| EC60.aa | (51) | TITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTD |
| EC61.aa | (51) | TITDYVTLQRGSAYGGVLSNFSGTVKEYSGSSYPFPTTSETPRVVYNSRTD |
| EC62.aa | (51) | TITDYVTLQRGSAYGGVLSHFSGTVKYSGSSYPFPTTSETPRVVYNSRTD |
| EC80.aa | (51) | TITDYVTLQRGSAYGGVLSFSFGTVKYNGSSYPFPTTSETPRVVYNSRTD |
| EC89.aa | (51) | TITDYVTLQRGSAYGGVLSFSFGTVKYNGSSYPFPTTSETPRVVYNSRTD |
| EC95.aa | (51) | TITDYVTLQRGSAYGGVLSFSFGTVKYNGSSYPFPTTSETPRVVYNSRTD |
| G189.aa | (51) | TITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTD |
| J96.aa | (51) | TITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTD |
| NU14.aa | (51) | TITDYVTLQRGAAYGGVLSFSFGTVKYNGSSYPFPTTSETPRVVYNSRTD |
| Consensus | (51) | TITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTD |

Figure 2(b)

| | | 101 | 150 |
|-----------|-------|------------|--|
| B210.aa | (101) | KPWPVALYLT | TPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| B212.aa | (101) | KPWPVALYLT | TPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| B217.aa | (101) | KPWPVALYLT | TPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| B223.aa | (101) | KPWPVALYLT | TPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| B228.aa | (101) | KPWPVALYLT | TPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| B238.aa | (101) | KPWPVALYLT | TPVSSAGGVVIAKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| B240.aa | (101) | KPWPVALYLT | TPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| B242.aa | (101) | KPWPVALYLT | TPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| DS17.aa | (101) | KPWPVALYLT | TPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| EC42.aa | (101) | KPWPVALYLT | TPVSSAGGVVIAKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| EC45.aa | (101) | KPWPVALYLT | TPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| EC56.aa | (101) | KPWPVALYLT | TPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| EC58.aa | (101) | KPWPVALYLT | TPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| EC60.aa | (101) | KPWPVALYLT | TPVSSAGGVVIAKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| EC61.aa | (101) | KPWPVALYLT | TPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| EC62.aa | (101) | KPWPVALYLT | TPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| EC80.aa | (101) | KPWPVALYLT | TPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| EC89.aa | (101) | KPWPVALYLT | TPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| EC95.aa | (101) | KPWPVALYLT | TVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| G189.aa | (101) | KPWPVALYLT | TPVSSAGGVAIKAGSLIAVLILRQTKNYNSDDFQFVWNIYA |
| J96.aa | (101) | KPWPVALYLT | TPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| NU14.aa | (101) | KPWPVALYLT | TPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |
| Consensus | (101) | KPWPVALYLT | TPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYA |

| | | 151 | 200 |
|-----------|-------|---------------------|---------------------------------|
| B210.aa | (151) | NNDVVVPTGGCDASARDVT | VTLPDYRGSVPIPLTVYCAKSQNLGYLLSGT |
| B212.aa | (151) | NNDVVVPTGGCDASARDVT | VTLPDYRGSVPIPLTVYCAKSQNLGYLLSGT |
| B217.aa | (151) | NNDVVVPTGGCDVSARDVT | VTLPDYRGSVPIPLTVYCAKSQNLGYLLSGT |
| B223.aa | (151) | NNDVVVPTGGCDVSARDVT | VTLPDYRGSVPIPLTVYCAKSQNLGYLLSGT |
| B228.aa | (151) | NNDVVVPTGGCDVSAHDVT | VTLPDYRGSVPIPLTVYCAKSQNLGYLLSGT |
| B238.aa | (151) | NNDVVVPTGGCDVSARDVT | VTLPDYRGSVPIPLTVYCAKSQNLGYLLSGT |
| B240.aa | (151) | NNDVVVPTGGCDVSARDVT | VTLPDYRGSVPIPLTVYCAKSQNLGYLLSGT |
| B242.aa | (151) | NNDVVVPTGGCDVSAHDVT | VTLPDYRGSVPIPLTVYCAKSQNLGYLLSGT |
| DS17.aa | (151) | NNDVVVPTGGCDASARDVT | VTLPDYRGSVPIPLTVYCAKSQNLGYLLSGT |
| EC42.aa | (151) | NNDVVVPTGGCDVSARDVT | VTLPDYPGSVPIPLTVYCAKSQNLGYLLSGT |
| EC45.aa | (151) | NNDVVVPTGGCDVSARDVT | VTLPDYPGSVPIPLTVYCAKSQNLGYLLSGT |
| EC56.aa | (151) | NNDVVVPTGGCDVSARDVT | VTLPDYPGSVPIPLTVYCAKSQNLGYLLSGT |
| EC58.aa | (151) | NNDVVVPTGGCDVSARDVT | VTLPDYRGSVPIPLTVYCAKSQNLGYLLSGT |
| EC60.aa | (151) | NNDVVVPTGGCDVSARDVT | VTLPDYRGSVPIPLTVYCAKSQNLGYLLSGT |
| EC61.aa | (151) | NNDVVVPTGGCDVSARDVT | VTLPDYRGSVPIPLTVYCAKSQNLGYLLSGT |
| EC62.aa | (151) | NNDVVVPTGGCDVSARDVT | VTLPDYRGSVPIPLTVYCAKSQNLGYLLSGT |
| EC80.aa | (151) | NNDVVVPTGGCDVSARDVT | VTLPDYRGSVPIPLTVYCAKSQNLGYLLSGT |
| EC89.aa | (151) | NNDVVVPTGGCDVSARDVT | VTLPDYRGSVPIPLTVYCAKSQNLGYLLSGT |
| EC95.aa | (151) | NNDVVVPTGGCDVSARDVT | VTLPDYRGSVPIPLTVYCAKSQNLGYLLSGT |
| G189.aa | (151) | NNDVVVPTGGCDVSARDVT | VTLPDYRGSVPIPLTVYCAKSQNLGYLLSGT |
| J96.aa | (151) | NNDVVVPTGGCDVSARDVT | VTLPDYPGSVPIPLTVYCAKSQNLGYLLSGT |
| NU14.aa | (151) | NNDVVVPTGGCDVSARDVT | VTLPDYRGSVPIPLTVYCAKSQNLGYLLSGT |
| Consensus | (151) | NNDVVVPTGGCDVSARDVT | VTLPDYRGSVPIPLTVYCAKSQNLGYLLSGT |

Figure 2(c)

| | 201 | 250 |
|-----------|-------|--|
| B210.aa | (201) | HADAGNSIFTNTASFSPAQGVGVQLARNGTVIPANNTVSLGAVGTSVAVSL |
| B212.aa | (201) | HADAGNSIFTNTASFSPAQGVGVQLTRNGTIIIPANNTVSLGAVGTSVAVSL |
| B217.aa | (201) | HADAGNSIFTNTASFSPAQGVGVQLTRNGTIIIPANNTVSLGAVGTSVAVSL |
| B223.aa | (201) | HADAGNSIFTNTASFSPAQGVGVQLTRNGTIIIPANNTVSLGAVGTSVAVSL |
| B228.aa | (201) | HADAGNSIFTNTASFSPAQGVGVQLTRNGTIIIPANNTVSLGAVGTSVAVSL |
| B238.aa | (201) | HADAGNSIFTNTASFSPAQGVGVQLTRNGTIIIPANNTVSLGAVGTSVAVSL |
| B240.aa | (201) | HADAGNSIFTNTASFSPAQGVGVQLTRNGTIIIPANNTVSLGAVGTSVAVSL |
| B242.aa | (201) | HADAGNSIFTNTASFSPAQGVGVQLTRNGTIIIPANNTVSLGAVGTSVAVSL |
| DS17.aa | (201) | HADAGNSIFTNTASFSPAQGVGVQLTRNGTIIIPANNTVSLGAVGTSVAVSL |
| EC42.aa | (201) | TADAGNSIFTNTASFSPAQGVGVQLTRNGTIIIPANNTVSLGAVGTSVAVSL |
| EC45.aa | (201) | TADAGNSIFTNTASFSPAQGVGVQLTRNGTIIIPANNTVSLGAVGTSVAVSL |
| EC56.aa | (201) | TADAGNSIFTNTASFSPAQGVGVQLTRNGTIIIPANNTVSLGAVGTSVAVSL |
| EC58.aa | (201) | HADAGNSIFTNTASFSPAQGVGVQLTRNGTIIIPANNTVSLGAVGTSVAVSL |
| EC60.aa | (201) | HADAGNSIFTNTASFSPAQGVGVQLTRNGTIIIPANNTVSLGAVGTSVAVSL |
| EC61.aa | (201) | HADAGNSIFTNTASFSPAQGVGVQLTRNGTIIIPANNTVSLGAVGTSVAVSL |
| EC62.aa | (201) | HADAGNSIFTNTASFSPAQGVGVQLTRNGTIIIPANNTVSLGAVGTSVAVSL |
| EC80.aa | (201) | HADAGNSIFTNTASFSPAQGVGVQLTANGTIVPANNTVSLGAVGTSVAVSL |
| EC89.aa | (201) | HADAGNSIFTNTASFSPAQGVGVQLTANGTIVPANNTVSLGAVGTSVAVSL |
| EC95.aa | (201) | HADAGNSIFTNTASFSPAQGVGVQLTRNGTIIIPANNTVSLGAVGTSVAVSL |
| G189.aa | (201) | HADAGNSIFTNTASFSPAQGVGVQLTRNGTIIIPANNTVSLGAVGTSVAVSL |
| J96.aa | (201) | TADAGNSIFTNTASFSPAQGVGVQLTRNGTIIIPANNTVSLGAVGTSVAVSL |
| NU14.aa | (201) | HADAGNSIFTNTASFSPAQGVGVQLTRNGTIIIPANNTVSLGAVGTSVAVSL |
| Consensus | (201) | HADAGNSIFTNTASFSPAQGVGVQLTRNGTIIIPANNTVSLGAVGTSVAVSL |

| | 251 | 279 |
|-----------|-------|--------------------------------|
| B210.aa | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |
| B212.aa | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |
| B217.aa | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |
| B223.aa | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |
| B228.aa | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |
| B238.aa | (251) | GLTANYARTGGQVTAGNVQSIIGATFVYQ |
| B240.aa | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |
| B242.aa | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |
| DS17.aa | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |
| EC42.aa | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |
| EC45.aa | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |
| EC56.aa | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |
| EC58.aa | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |
| EC60.aa | (251) | GLTANYARTGGQVTAGNVRSIIIAVTFVYQ |
| EC61.aa | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |
| EC62.aa | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |
| EC80.aa | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |
| EC89.aa | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |
| EC95.aa | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |
| G189.aa | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |
| J96.aa | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |
| NU14.aa | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |
| Consensus | (251) | GLTANYARTGGQVTAGNVQSIIGVTFVYQ |

Fig. 3

Step1: Construction of pCGA101-8

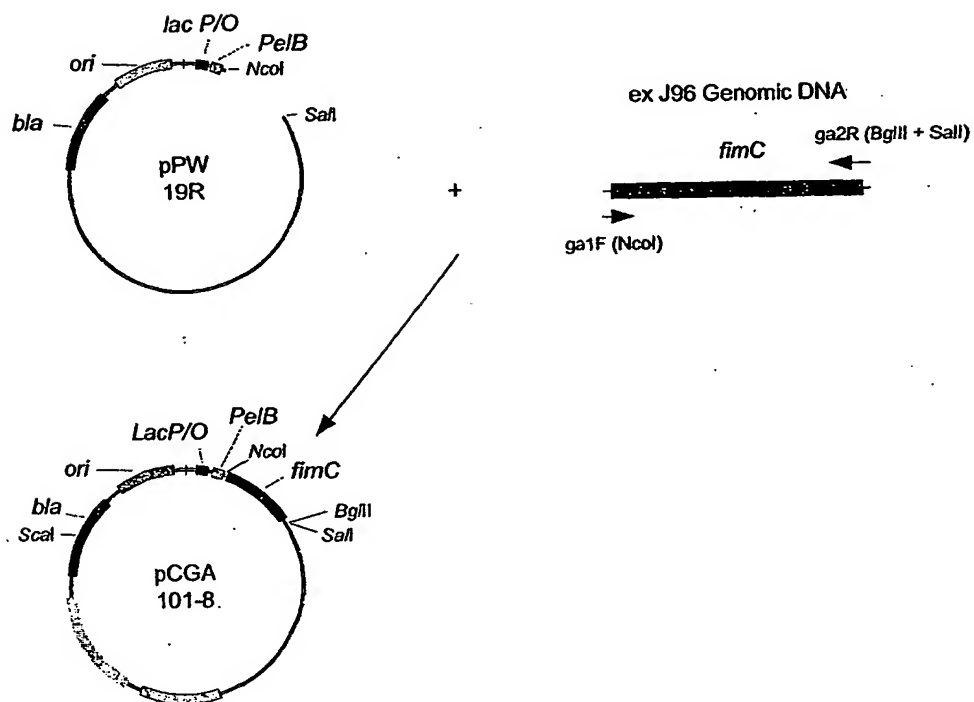
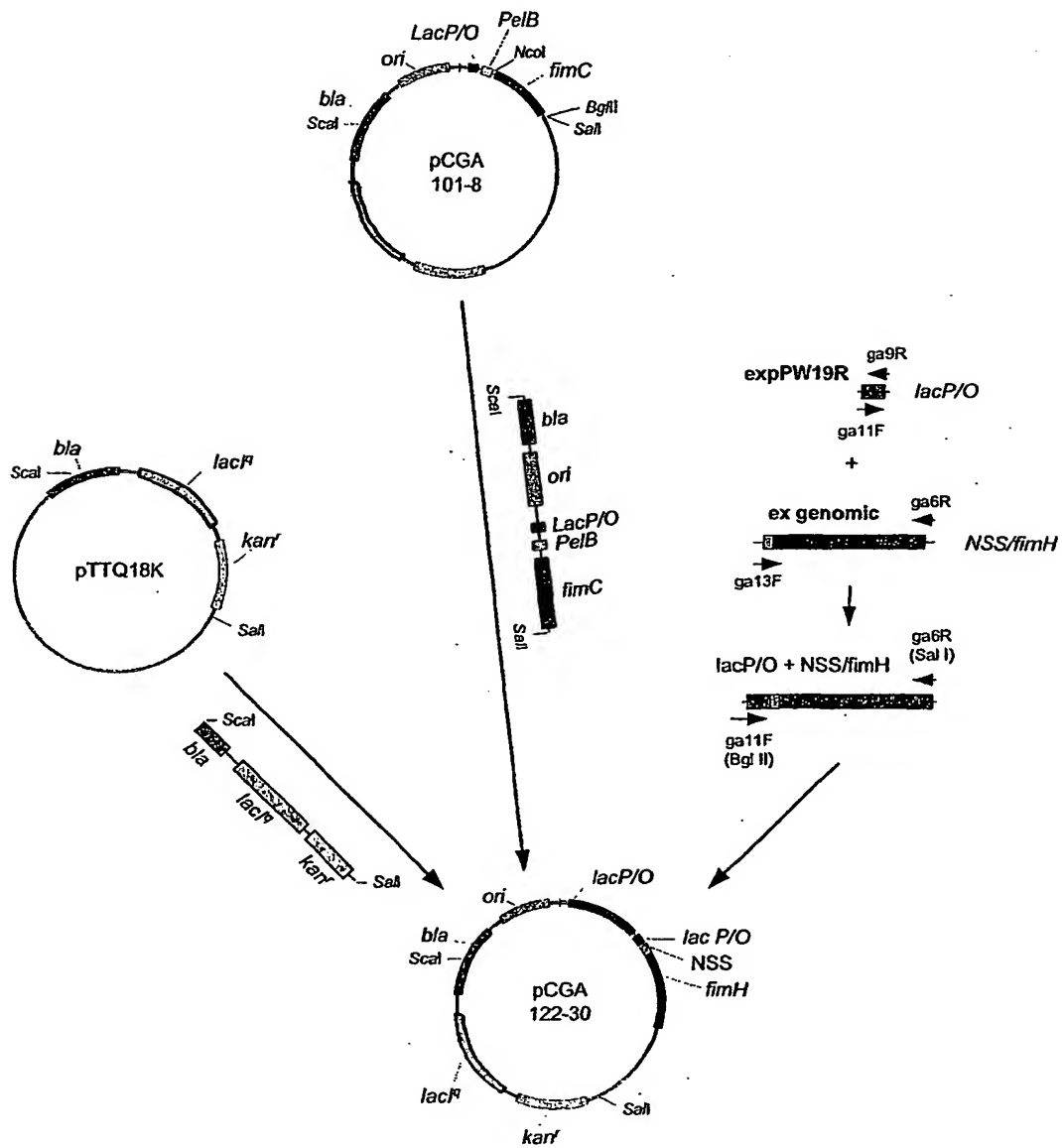


Fig. 4

Step 2: Construction of pCGA122-30



Step 3: Selection of final clone

Figs.

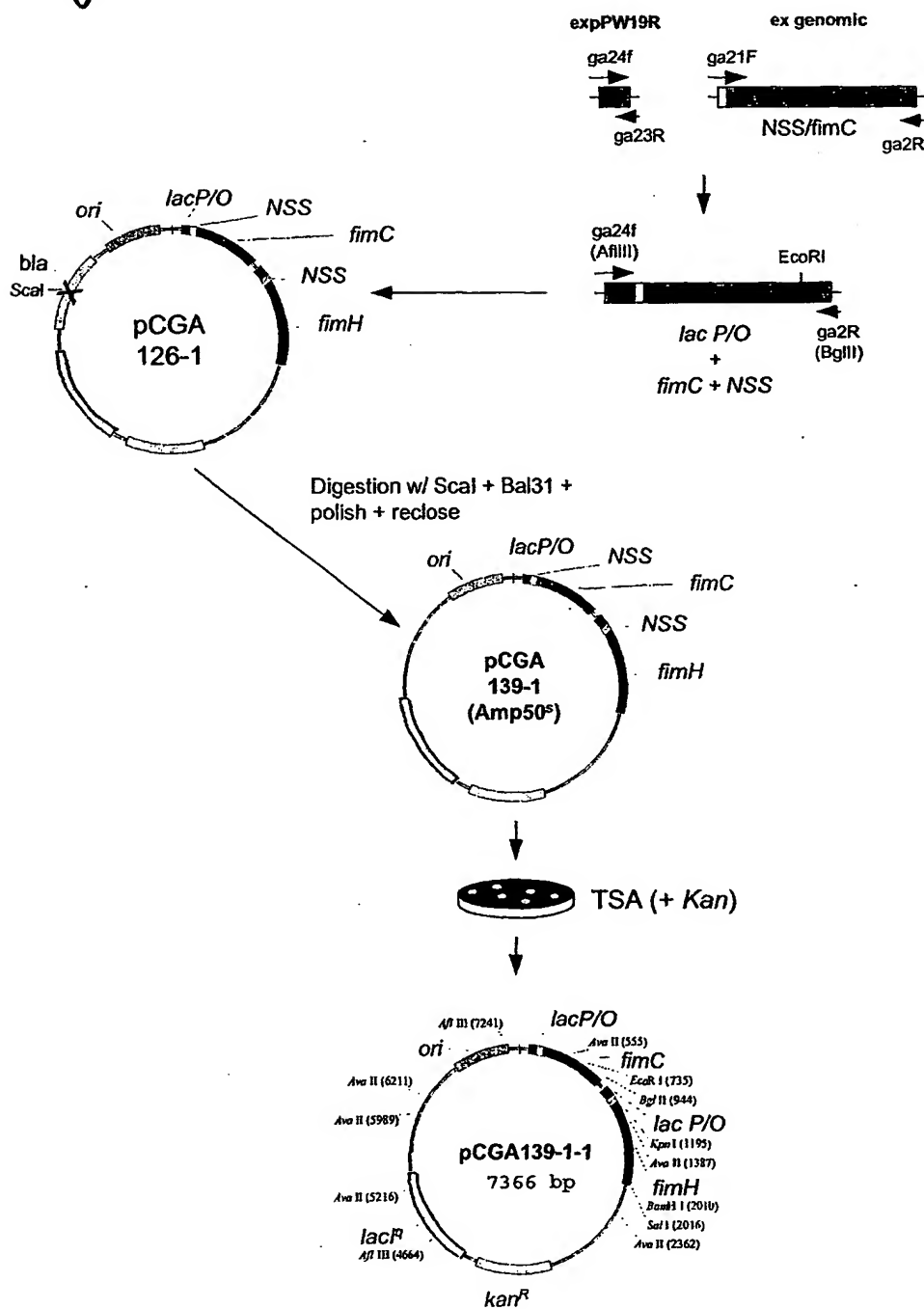
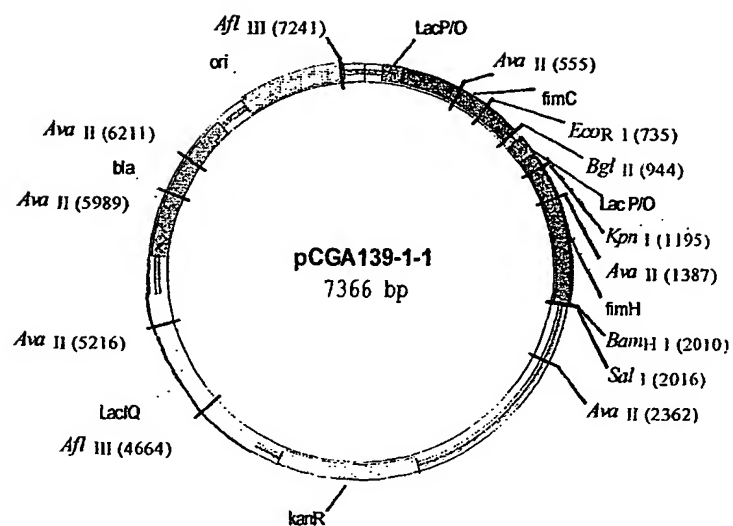


Fig. 6



SEQUENCE LISTING

<110> Langermann, Solomon
Revel, Andrew
Auguste, Christine
Burlein, Jeanne

<120> FimH Adhesin Proteins and Methods of Use

<130> 469201-549

<150> US/60/216,750

<151> 2000-07-07

<160> 64

<170> PatentIn version 3.0

<210> 1

<211> 837

<212> DNA

<213> E. coli

<400> 1

| | | | | | | |
|-------------|------------|-------------|------------|-------------|-------------|-----|
| ttcgccgtgta | aaaccgccaa | tggtaccgct | atccctattg | gcggtggcag | cgccaatggt | 60 |
| tatgtaaacc | ttgcgccgt | cgtgaatgtg | gggcaaaacc | tggtcgtgga | tctttcgacg | 120 |
| caaactcttt | gccataacga | ttatccggaa | accattacag | actatgtcac | actgcaacga | 180 |
| ggctcggctt | atggcggcgt | gttatctaata | ttttccggga | tcgtaaaata | tagtggcagt | 240 |
| agctatcctt | tccctaccac | cagcgaaacg | ccgcgcgttg | tttataattc | gagaacggat | 300 |
| aagccgtggc | cggtggcgct | ttatttgacg | cctgtgagca | gtgcgggggg | agtggcgatt | 360 |
| aaagcaggct | cattaattgc | cgtgcttatt | ttgcgacaga | ccaacaacta | taacagcgat | 420 |
| ggtttccagt | ttgtgtggaa | tatttacgcc | aataatgatg | tggtggtgcc | cactggcggc | 480 |
| tgcgatgctt | ctgctcgtga | tgtcacccgt | actctgccgg | actaccctgg | ttcagtgccg | 540 |
| attcctctta | ccgtttattg | tgcgaaaagc | caaaacctgg | ggtattacct | ctccggcaca | 600 |
| accgcagggtg | cgggcaactc | gattttcacc | aatacccgct | cgttttcacc | cgcgaggggc | 660 |
| gtcggcgtag | agttggcgcg | caacggtaacg | gttattccag | cgaataaacac | ggtatcggtta | 720 |
| ggagcagtag | ggacttcggc | ggtgagctctg | ggattaacgg | caaattacgc | acgtaccgga | 780 |
| gggcagggtga | ctgcaggga | tgtgcaatcg | attattggcg | tgacttttgt | ttatcaa | 837 |

<210> 2

<211> 837

<212> DNA

<213> E. coli

<400> 2

| | | | | | | |
|-------------|-------------|------------|------------|------------|------------|-----|
| ttcgccgtgta | aaaccgccaa | tggtaccgct | atccctattg | gcggtggcag | cgccaatggt | 60 |
| tatgtaaacc | ttgcgccgtgc | cgtgaatgtg | gggcaaaacc | tggtcgtgga | tctttcgacg | 120 |
| caaactcttt | gccataacga | ttaccggaa | accattacag | actatgtcac | actgcaacga | 180 |
| ggttcggctt | atggcggcgt | gttatctagt | ttttccggga | tcgtaaaata | taatggcagt | 240 |
| agctatcctt | tccctactac | cagcgaaacg | ccgcgggttg | tttataattc | gagaacggat | 300 |
| aagccgtggc | cggtggcgct | ttatttgacg | cctgtgagca | gtgcgggggg | agtggcgatt | 360 |
| aaagcaggct | cattaattgc | cgtgcttatt | ttgcgacaga | ccaacaacta | taacagcgat | 420 |

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| gatttccagt | ttgtgtggaa | tatttacgcc | aataatgatg | tggtggtgcc | cactggcggc | 480 |
| tgcgatgctt | ctgctcgtga | tgtcacggtt | actctgccgg | actaccctgg | ttcagtgccg | 540 |
| attcctctta | ccgtttattg | tgcgaaaagc | caaaacctgg | ggtattacct | ctccggcaca | 600 |
| accgcagatg | cgggcaactc | gattttcacc | aataccgcgt | cgttttcacc | cgcgcagggc | 660 |
| gtcggcgtag | agttgacgcy | caacggtagc | attattccag | cgaataacac | ggtatcgtaa | 720 |
| ggagcagtag | ggacttcggc | ggtaagtctg | ggattaacgg | caaattacgc | acgtaccgga | 780 |
| gggcaggtga | ctgcagggaa | tgtgcaatcg | attattggcg | tgacttttgt | ttatcaa | 837 |

<210> 3
 <211> 837
 <212> DNA
 <213> E. coli

| | | | | | | |
|------------|------------|-------------|------------|------------|------------|-----|
| <400> 3 | | | | | | |
| ttcgctgta | aaaccgcca | tggtacagct | atccctattg | gcggtggcag | cgctaattgt | 60 |
| tatgtaaacc | ttgcgcctgc | cgtgaatgtg | gggcaaaacc | tggtcgtaga | tctttcgacg | 120 |
| caaatctttt | gccataacga | ttatccggaa | accattacag | actatgtcac | actgcaacga | 180 |
| ggctcggcct | atggcggcgt | gttatctaat | ttttccggga | ccgtaaaata | tagtggcagt | 240 |
| agctatccat | ttccgaccac | cagcgaaaacg | ccgcgggttg | tttataattc | gagaacggat | 300 |
| aagccgtggc | cggtagcgct | ttatttgacg | cctgtgagca | gtgcggggcg | ggtggcgatt | 360 |
| aaagctggct | cattaattgc | cgtgcttatt | ttgcgacaga | ccaacaacta | taacagcgat | 420 |
| gatttccagt | ttgtgtggaa | tatttacgcc | aataatgatg | tggtggtgcc | tactggcggc | 480 |
| tgcgatgttt | ctgctcgtga | tgtcacggtt | actctgccgg | actaccctgg | ttcagtgcca | 540 |
| attcctctta | ccgtttattg | tgcgaaaagc | caaaacctgg | ggtattacct | ctccggcaca | 600 |
| accgcagatg | cgggcaactc | gattttcacc | aataccgcgt | cgttttcacc | agcgcagggc | 660 |
| gtcggcgtag | agttgacgcy | caacggtagc | attattccca | cgaataacac | ggtatcgtaa | 720 |
| ggagcagtag | ggacttcggc | ggtaagtctg | ggattaacgg | caaattacgc | acgtaccgga | 780 |
| gggcaggtga | ctgcagggaa | tgtgcaatcg | attattggcg | tgacttttgt | ttatcaa | 837 |

<210> 4
 <211> 840
 <212> DNA
 <213> E. coli

| | | | | | | |
|------------|------------|-------------|------------|------------|------------|-----|
| <400> 4 | | | | | | |
| ttcgctgta | aaaccgcca | tggtaccgca | atccctattg | gcggtggcag | cgccaatgtt | 60 |
| tatgtaaacc | ttgcgcctgc | cgtgaatgtg | gggcaaaacc | tggtcgtaga | tctttcgacg | 120 |
| caaatctttt | gccataacga | ttaccagaa | accattacag | actatgtcac | actgcaacga | 180 |
| ggtgcggtct | atggcggcgt | gttatctagt | ttttccggga | ccgtaaaata | taatggcagt | 240 |
| agctatcctt | tccctactac | cagcgaaaacg | ccgcgggttg | tttataattc | gagaacggat | 300 |
| aagccgtggc | cggtagcgct | ttatttgacg | ccggtgagca | gtgcgggggg | agtggcgatt | 360 |
| aaagctggct | cattaattgc | cgtgcttatt | ttgcgacaga | ccaacaacta | taacagcgat | 420 |
| gatttccagt | ttgtgtggaa | tatttacgcc | aataatgatg | tggtggtgcc | cactggcggc | 480 |
| tgcgatgttt | ctgctcgtga | tgtcacggtt | actctgccgg | actaccctgg | ttcagtgccg | 540 |
| attcctctta | ccgtttattg | tgcgaaaagc | caaaacctgg | ggtattacct | ctccggcaca | 600 |
| accgcagatg | cgggcaactc | gattttcacc | aataccgcgt | cgttttcacc | cgcgcagggc | 660 |
| gtcggcgtag | agttgacgcy | caacggtagc | attattccag | cgaataacac | ggtatcgtaa | 720 |
| ggagcagtag | ggacttcggc | ggtaagtctg | ggattaacgg | caaattacgc | acgtaccgga | 780 |
| gggcaggtga | ctgcagggaa | tgtgcaatcg | attattggcg | tgacttttgt | ttatcaataa | 840 |

<210> 5

<211> 840
 <212> DNA
 <213> E. coli

<400> 5
 ttgcctgta aaaccgcaa tggtagcgt attcctattg gcggtggcag cgctaagtgt 60
 tatgtaaacc ttgcgcctgc cgtgaatgtg gggcaaaacc tggtcgtaga tctttcgacg 120
 caaatctttt gccataacga ttatccggaa accattacag actatgtcac actgcaacga 180
 ggctcggcct atggcggcgt gttatctaatt tttccggga ccgtaaaata tagtggcagt 240
 agctatccat ttccgactac cagcgaaaacg ccgcgggttg tttataattc gagaacggat 300
 aagccgtggc cgggtggcgt ttatttgacg cctgtgagca gtgcgggtgg ggtggcgatt 360
 aaagctggct cattaattgc cgtgcttatt ttgcgacaga ccaacaacta taacagcgat 420
 gatttccagt ttgtgtggaa tatttacgcc aataatgatg tgggtggtgcc tactggcggc 480
 tgcgatgttt ctgctcatga tgtcaccgtt actctgccgg actaccctgg ttcagtgcc 540
 attcctctta ccgtttattg tgcgaaaagc caaaacctgg ggtattacct ctccggcaca 600
 accgcagatg cgggcaactc gattttcacc aataccgcgt cgttttcacc agcgcgaggc 660
 gtcggcgtag agttgacgag caacggtagc attattccag cgaataaacac ggtatcgtaa 720
 ggagcagtag ggacttcggc ggtaagtctg ggattaacgc caaattacgc acgtaccgga 780
 gggcaggtga ctgcagggaa tgtgcaatcg attattggcg tgacttttgt ttatcaataa 840

<210> 6
 <211> 837
 <212> DNA
 <213> E. coli

<400> 6
 ttgcctgta aaaccgcaa tggcaccgt atccctattg gcggtggcag cgccaatgtt 60
 tatgtaaaca ttgcgcccgc cgtgaatgtg gggcaaaacc tggtcgtgga tctttcgacg 120
 caaatctttt gccataacga ttaccggaa accattacag attatgtcac actgcaacga 180
 ggctcggcct atggcggcgt gttatctaatt tttccggga ccgtaaaata tagtggcagt 240
 agctatccat ttccgaccac cagtgaacgc ccgcgggttg tttataattc gagaacggat 300
 aagccgtggc cgggtggcgt ttatttgacg cctgtgagca gtgcggggcg ggtgggtgatt 360
 aaagctggct cattaattgc cgtgcttatt ttgcgacaga ccaacaacta taacagcgat 420
 gatttccagt ttgtgtggaa tatttacgcc aataatgatg tgggtggtgcc cactggcggc 480
 tgcgatgttt ctgctcgtga tgtcaccgtt actctgccgg actaccctgg ttcagtgcc 540
 attcctctta ccgtttattg tgcgaaaagc caaaacctgg ggtattacct ctccggcaca 600
 accgcagatg cgggcaactc gattttcacc aataccgcgt cgttttcacc tgcacagggc 660
 gtcggcgtag agttgacgag caacggtagc attattccag cgaataaacac ggtatcgtaa 720
 ggagcagtag ggacttcggc ggtaagtctg ggattaacgc caaattacgc acgtaccgga 780
 gggcaggtga ctgcagggaa tgtgcaatcg attattgccg tgacttttgt ttatcaa 837

<210> 7
 <211> 837
 <212> DNA
 <213> E. coli

<400> 7
 ttgcctgta aaaccgcaa tggtagcgt atccctattg gcggtggcag cgccaatgtt 60
 tatgtaaacc ttgcgcccgt cgtgaatgtg gggcaaaacc tggtcgtgga tctttcgacg 120
 caaatctttt gccataacga ttatccggaa accattacag actatgtcac actgcaacga 180
 ggctcggcct atggcggcgt gttatctaatt tttccggga ccgtaaaata tagtggcagt 240
 agctatccat ttccctaccac cagcgaaaacg ccgcgcgttg tttataattc gagaacggat 300


```

aagccgtggc cgggtggcgct ttatttgacg cctgtgagca gtgcggggcg gttggtgatt 360
aaagctggct cattaattgc cgtgcttatt ttgcgacaga ccaacaacta taacagcgat 420
gatttccagt ttgtgtggaa tatttacgcc aataatgatg tgggtggtgcc tactggcggc 480
tgcatgttt ctgctcgtga tgtcacgctt actctgccgg actaccgtgg ttcagtgccca 540
attcctctta ccgtttattg tgcgaaaagc caaaacctgg ggtattacct ctccggcaca 600
accgcagatg cgggcaactc gattttcacc aataccgcgt cgttttcacc tgcacagggc 660
gtcggcgtag agttgacgcg caacggtagc attattccaa cgaataacac ggtatcgtaa 720
ggagcagtag ggacttcggc ggtaagtctg ggattaacgg caaattacgc acgtaccgga 780
gggcaggtga ctgcagggaa tgtgcaatcg attattggcg tgacttttgt ttatcaa 837

```

```

<210> 8
<211> 837
<212> DNA
<213> E. coli

```

```

<400> 8
tttgctgta aaaccgccaa tggcacgct atccctattg gcggtggcag cgccaatgtt 60
tatgtaaact tggcgcccg cgtgaatgtg gggcaaaacc tggctcgtag tctttcgacg 120
caaacctttt gccataacga ttatccggaa accattacag actatgtcac actgcaacga 180
ggctcggtct atggcgcgct gttatctaat tttccggga ccgtaaaata tagtggcagt 240
agctatccat tcccgactac cagcgaaacg ccgcggttg ttataattc gagaacggat 300
aagccgtggc cgggtggcgct ttatttgacg cctgtgagca gtgcgggtgg ggtggcgatt 360
aaagctggct cattaattgc cgtgcttatt ttgcgacaga ccaacaacta taacagcgat 420
gatttccagt ttgtgtggaa tatttacgcc aataatgatg tgggtggtgcc tactggcggc 480
tgcatgttt ctgctcatga tgtcacgctt actctgccgg actaccctgg ttcagtgccca 540
attcctctta ccgtttattg tgcgaaaagc caaaacctgg ggtattacct ctccggcaca 600
accgcagatg cgggcaactc gattttcacc aataccgcgt cgttttcacc agcgagggc 660
gtcggcgtag agttgacgcg caacggtagc attattccag cgaataacac ggtatcgtaa 720
ggagcagtag ggacttcggc ggtgagtctg ggattaacgg caaattacgc acgtaccgga 780
gggcaggtga ctgcagggaa tgtgcaatcg attattggcg tgacttttgt ttatcaa 837

```

```

<210> 9
<211> 837
<212> DNA
<213> E. coli

```

```

<400> 9
ttcgctgta aaaccgccaa tggtagcgca atccctattg gcggtggcag cgccaatgtt 60
tatgtaaacc ttgcgctgc cgtgaatgtg gggcaaaacc tggctgtaga tctttcgacg 120
caaactttt gccataacga ttaccagaa accattacag actatgtcac actgcaacga 180
ggttcggtct atggcgcgct gttatctagt tttccggga ccgtaaaata taatggcagt 240
agctatcctt tccctactac cagcgaaacg ccgcggttg ttataattc gagaacggat 300
aagccgtggc cgggtggcgct ttatttgacg ccggtgagca gtgcgggggg agtggcgatt 360
aaagctggct cattaattgc cgtgcttatt ttgcgacaga ccaacaacta taacagcgat 420
gatttccagt ttgtgtggaa tatttacgcc aataatgatg tgggtggtgcc cactggcggc 480
tgtgatgctt ctgctcgtga tgtcacgctt actttgccgg actaccctgg ttcagtgccg 540
attcctctta ccgtttattg tgcgaaaagc caaaacctgg ggtattacct atccggcaca 600
accgcagatg cgggcaactc gattttcacc aataccgcgt cgttttcacc cgcgagggc 660
gtcggcgtag agttgacgcg caacggtagc attattccag cgaataacac ggtatcgtaa 720
ggagcagtag ggacttcggc ggtaagtctg ggattaacgg caaattacgc acgtaccgga 780
gggcaggtga ctgcagggaa tgtgcaatcg attattggcg tgacttttgt ttatcaa 837

```

<210> 10
 <211> 840
 <212> DNA
 <213> E. coli

<400> 10
 ttgcctgta aaaccgcaa tggcaccgct atccctattg gcggtggcag cgccaatgtt 60
 tatgtaaacc ttgcgcccgc cgtgaatgtg gggcaaaacc tggtcgtgga tctttcgacg 120
 caaatctttt gccataacga ttaccggaa accattacag attatgtcac actgcaacga 180
 ggctcggtt atggcgccgt gttatctaatt tttccggga ccgtaaaata tagtggcagt 240
 agctatccat ttccgaccac cagtgaacg ccgcgggttg tttataattc gagaacggat 300
 aagccgtggc cgtggcgct ttatttgacg cctgtgagca gtgcggggcg ggtggtgatt 360
 aaagctggct cattaattgc cgtgcttatt ttgcgacaga ccaacaacta taacagcgat 420
 gatttccagt ttgtgtggaa tatttacgcc aataatgatg tgggtgtgcc cactggcgcc 480
 tgcgatgttt ctgctcgtga tgtcaccgtt actctgccg actaccctgg ttcagtgcg 540
 attcctctta ccgtttattg tgcgaaaagc caaaacctgg ggtattacct ctccggcaca 600
 accgcagatg cgggcaactc gattttcacc aataccgcgt cgttttcacc tgcacagggc 660
 gtcggcgtag agttgacgcg caacggtacg attattccag cgaataacac ggtatcgta 720
 ggagcagtag ggacttcggc ggtaagtctg ggattaacg caaattacgc acgtaccgga 780
 gggcaggtga ctgcaggga tgtgcaatcg attattggcg tgacttttgt ttatcaataa 840

<210> 11
 <211> 837
 <212> DNA
 <213> E. coli

<400> 11
 ttgcctgta aaaccgcaa tggtaaccga atccctattg gcggtggcag cgccaatgtt 60
 tatgtaaacc ttgcgcctgc cgtgaatgtg gggcaaaacc tggtcgtaga tctttcgacg 120
 caaatctttt gccataacga ttaccagaa accattacag actatgtcac actgcaacga 180
 ggtgcggctt atggcgccgt gttatctagt tttccggga ccgtaaaata taatggcagt 240
 agctatcctt tccctactac cagcgaaacg ccgcgggttg tttataattc gagaacggat 300
 aagccgtggc cgtggcgct ttatttgacg ccggtgagca gtgcggggcg agtggcgatt 360
 aaagctggct cattaattgc cgtgcttatt ttgcgacaga ccaacaacta taacagcgat 420
 gatttccagt ttgtgtggaa tatttacgcc aataatgatg tgggtgtgcc cactggcgcc 480
 tgcgatgttt ctgctcgtga tgtcaccgtt actctgccg actaccctgg ttcagtgcg 540
 attcctctta ccgtttattg tgcgaaaagc caaaacctgg ggtattacct ctccggcaca 600
 accgcagatg cgggcaactc gattttcacc aataccgcgt cgttttcacc cgcgcagggc 660
 gtcggcgtag agttgacgcg caacggtacg attattccag cgaataacac ggtatcgta 720
 ggagcagtag ggacttcggc ggtaagtctg ggattaacg caaattacgc acgtaccgga 780
 gggcaggtga ctgcaggga tgtgcaatcg attattggcg tgacttttgt ttatcaa 837

<210> 12
 <211> 840
 <212> DNA
 <213> E. coli

<400> 12
 ttgcctgta aaaccgcaa tggtaaccgt atccctattg gcggtggcag cgccaatgtt 60
 tatgtaaacc ttgcgcccgc cgtgaatgtg gggcaaaacc tggtcgtgga tctttcgacg 120
 caaatctttt gccataacga ttaccggaa accattacag actatgtcac actgcaacga 180

| | | | | | | |
|------------|------------|-------------|------------|------------|------------|-----|
| ggctcggtt | atggcggtt | gttatcta | ttttccgga | ccgtaaaata | tagtggcagt | 240 |
| agctatccat | ttcctaccac | cagcgaaaacg | ccgcgcgtt | tttataattc | gagaacggat | 300 |
| aagccgtggc | cggtggcgct | ttatttgacg | cctgtgagca | gtgcggggcg | ggtggcgatt | 360 |
| aaagctggct | cattaattgc | cgtgcttatt | ttgcgacaga | ccaacaacta | taacagcgat | 420 |
| gatttccagt | ttgtgtggaa | tatttacgcc | aataatgatg | tggtggtgcc | tactggcggc | 480 |
| tgcgatgttt | ctgctcgtga | tgtcaccggt | actctgccgg | actaccctgg | ttcagtgcc | 540 |
| attcctctta | ccgtttattg | tgcgaaaagc | caaaacctgg | ggtattacct | ctccggcaca | 600 |
| accgcagatg | cgggcaactc | gattttcacc | aataccgcgt | cgttttcacc | tgcacagggc | 660 |
| gtcggcgtag | agttgacg | caacggtagc | attattccag | cgaataacac | ggtatcgta | 720 |
| ggagcagtag | ggacttcggc | ggtgagctcg | ggattaacgg | caaattatgc | acgtaccgga | 780 |
| gggcaggtga | ctgcagggaa | tgtgcaatcg | attattggcg | tgacttttgt | ttatcaataa | 840 |

<210> 13
 <211> 837
 <212> DNA
 <213> E. coli

| | | | | | | |
|------------|------------|-------------|------------|------------|------------|-----|
| <400> 13 | | | | | | |
| ttcgctgta | aaaccgccaa | tggtaccgca | atccctattg | gcggtggcag | cgccaatggt | 60 |
| tatgtaaacc | ttgcgcctgc | cgtgaatgtg | gggcaaaacc | tggtcgtaga | tctttcgacg | 120 |
| caaactcttt | gccataacga | ttaccagaa | accattacag | actatgtcac | actgcaacga | 180 |
| ggttcggctt | atggcagcgt | gttatctagt | ttttccgga | ccgtaaaata | taatggcagt | 240 |
| agctatcctt | tcctactac | cagcgaaaacg | ccgcgggtt | tttataattc | gagaacggat | 300 |
| aagccgtggc | cggtggcgct | ttatttgacg | ccggtgagca | gtgcgggggg | agtggcgatt | 360 |
| aaagctggct | cattaattgc | cgtgcttatt | ttgcgacaga | ccaacaacta | taacagcgat | 420 |
| gatttccagt | ttgtgtggaa | tatttacgcc | aataatgatg | tggtggtgcc | cactggcggc | 480 |
| tgtgatgttt | ctgctcgtga | tgtcaccggt | actctgccgg | actaccctgg | ttcagtgcc | 540 |
| attcctctta | ccgtttattg | tgcgaaaagc | caaaacctgg | ggtattacct | atccggcaca | 600 |
| accgcagatg | cgggcaactc | gattttcacc | aataccgcgt | cgttttcacc | cgcgagggc | 660 |
| gtcggcgtag | agttgacg | caacggtagc | attattccag | cgaataacac | ggtatcgta | 720 |
| ggagcagtag | ggacttcggc | ggtgagctcg | ggattaacgg | caaattacgc | acgtaccgga | 780 |
| gggcaggtga | ctgcagggaa | tgtgcaatcg | attattggcg | tgacttttgt | ttatcaa | 837 |

<210> 14
 <211> 837
 <212> DNA
 <213> E. coli

| | | | | | | |
|------------|------------|-------------|------------|------------|------------|-----|
| <400> 14 | | | | | | |
| ttcgctgta | aaaccgccaa | tggtaccgca | atccctattg | gcggtggcag | cgccaatggt | 60 |
| tatgtaaacc | ttgcgcctgc | cgtgaatgtg | gggcaaaacc | tggtcgtaga | tctttcgacg | 120 |
| caaactcttt | gccataacga | ttaccagaa | accattacag | actatgtcac | actgcaacga | 180 |
| ggttcggctt | atggcagcgt | gttatctagt | ttttccgga | ccgtaaaata | taatggcagt | 240 |
| agctatcctt | tcctactac | cagcgaaaacg | ccgcgggtt | tttataattc | gagaacggat | 300 |
| aagccgtggc | cggtggcgct | ttatttgacg | ccggtgagca | gtgcgggggg | agtggcgatt | 360 |
| aaagctggct | cattaattgc | cgtgcttatt | ttgcgacaga | ccaacaacta | taacagcgat | 420 |
| gatttccagt | ttgtgtggaa | tatttacgcc | aataatgatg | tggtggtgcc | cactggcggc | 480 |
| tgtgatgttt | ctgctcgtga | tgtcaccggt | actctgccgg | actaccctgg | ttcagtgcc | 540 |
| attcctctta | ccgtttattg | tgcgaaaagc | caaaacctgg | ggtattacct | atccggcaca | 600 |
| accgcagatg | cgggcaactc | gattttcacc | aataccgcgt | cgttttcacc | cgcgagggc | 660 |
| gtcggcgtag | agttgacg | caacggtagc | attattccag | cgaataacac | ggtatcgta | 720 |
| ggagcagtag | ggacttcggc | ggtgagctcg | ggattaacgg | caaattacgc | acgtaccgga | 780 |

gggcagggtga ctgcagggaa tgtgcaatcg attattggcg tgacttttgt ttatcaa 837

<210> 15
 <211> 837
 <212> DNA
 <213> E. coli

<400> 15
 ttgcctata aaaccgccaa tggtagcgct atccctattg gcggtggcag cgccaatggt 60
 tatgtaaacc ttgcgccgc cgtgaatgtg gggcaaaacc tggtagtga tctttcgacg 120
 caaatctttt gccataacga ttatccggaa accattacag actatgtcac actgcaacga 180
 ggctcggctt atggcggcgt gttatctaat ttttcggga ccgtagaata tagtggcagt 240
 agctatccat ttcctaccac cagcgaaacg ccgcgcgttg tttataattc gagaacggat 300
 aagccgtggc cggtagcgct ttatttgacg cctgtgagca gtgcggggcg ggtggcgatt 360
 aaagctggct cattaattgc cgtgcttatt ttgcgacaga ccaacaacta taacagcgat 420
 gatttccagt ttgtgtggaa tatttacgcc aataatgatg tggtagtgcc tactggcggc 480
 tgcgatgttt ctgctcgtga tgtcacgctt actctgccgg actaccctgg ttcagtgcga 540
 attcctctta ccgtttattg tgcgaaaagc caaacctgg ggtattacct ctccggcaca 600
 accgcagatg cgggcaactc gattttcacc aataccgcgt cgttttcacc tgcacagggc 660
 gtcggcgtag agttgacgcg caacggtagc attattccag cgaataacac ggtatcttta 720
 ggagcagtag ggacttcggc ggtgagctcg ggattaacgg caaattatgc acgtaccgga 780
 gggcagggtga ctgcagggaa tgtgcaatcg attattggcg tgacttttgt ttatcaa 837

<210> 16
 <211> 837
 <212> DNA
 <213> E. coli

<400> 16
 atcgctgta aaaccgccaa tggcacgctt atccctattg gcggtggcag cgccaatggt 60
 tatgtaaacc ttgcgccgc cgtgaatgtg gggcaaaacc tggtagtga tctttcgacg 120
 caaatctttt gccataacga ttaccggaa accattacag actatgtcac actgcaacga 180
 ggctcggctt atggcggcgt gttatctcat ttttcggga ccgtaaaata tagtggcagt 240
 agctatccat ttcctaccac cagcgaaacg ccgcgcgttg tttataattc gagaacggat 300
 aagccgtggc cggtagcgct ttatttgacg cctgtgagca gtgcgggtgg ggtggcgatt 360
 aaggctggct cattaatggc tgtgctaatt ttgcgacaga ccaataacta taacagcgat 420
 gatttccagt ttgtgtggaa tatttacgcc aataatgatg tggtagtgcc cactggcggc 480
 tgtgatgttt ctgctcgtga tgtcacgctt actctgccag actaccctgg ttcagtgcga 540
 attcctctta ccgtttattg tgcgaaaagc caaacctgg ggtattacct ctccggcaca 600
 accgcagatg cgggcaactc gattttcacc aataccgcgt cgttttcacc tgcacagggc 660
 gtcggcgtag agttaacgcg caacggtagc attaatccag cgaataacac ggtatcgta 720
 ggagcagtag ggacttcggc ggtaagtctg ggattaacgg caaattacgc acgtaccgga 780
 gggcagggtga ctgcagggaa tgtgcaatcg attattggcg tgacttttgt ttatcaa 837

<210> 17
 <211> 837
 <212> DNA
 <213> E. coli

<400> 17
 ttgcctgta aaaccgccaa tggtagcgct atccctattg gcggtggcag cgccaatggt 60

| | | | | | | |
|------------|------------|------------|------------|-------------|-------------|-----|
| tatgtaaacc | ttgcgcccgt | cgtgaatgtg | gggcaaaacc | tggtcgtgga | tctttcgacg | 120 |
| caaatctttt | gccataacga | ttatccggaa | accattacag | actatgtcac | actgcaacga | 180 |
| ggctcggcct | atggcgccgt | gttatctaat | ttttccggga | cgtaaaaata | tagtggcagt | 240 |
| agctatccat | ttcctaccac | cagcgaaacg | cgcgcgttg | tttataattc | gagaacggat | 300 |
| aagccgtggc | cgggtggcgt | ttatgtgacg | cctgtgagca | gtgcggggcg | ggtggcgatt | 360 |
| aaagctggct | cattaattgc | cgtgcttatt | ttgcgacaga | ccaacaacta | taacagcgat | 420 |
| gatttccagt | ttgtgtggaa | tatttacgcc | aataatgatg | tggtgggtgcc | tactggcggc | 480 |
| tgcgatgttt | ctgctcgtga | tgtcacccgt | actctgccgg | actaccgtgg | ttcagtgccca | 540 |
| attcctctta | ccgtttattg | tgcgaaaagc | caaaacctgg | ggtattacct | ctccggcaca | 600 |
| cacgcagatg | cgggcaactc | gattttcacc | aataccgcgt | cgttttcacc | tgcacagggc | 660 |
| gtcggcgtag | agttgacgcg | caacggtagc | attattccag | cgaataacac | ggtatcggtta | 720 |
| ggagcagtag | ggacttcggc | ggtgagtcgt | ggattaacgg | caaattatgc | acgtaccgga | 780 |
| gggcaggtga | ctgcagggaa | tgtgcaatcg | attattggcg | tgacttttgt | ttatcaa | 837 |

<210> 18
 <211> 837
 <212> DNA
 <213> E. coli

| | | | | | | |
|------------|------------|-------------|------------|-------------|-------------|-----|
| <400> 18 | | | | | | |
| ttcgcctgta | aaaccgccaa | tggtaccgct | atccctattg | gcggtggcag | cgccaatggt | 60 |
| tatgtaaacc | ttgcgcctgc | cgtgaatgtg | gggcaaaacc | tggtcgtgga | tctttcgacg | 120 |
| caaatctttt | gccataacga | ttaccgggaa | accattacag | actatgtcac | actgcaacga | 180 |
| ggttcggcct | atggcgccgt | gttatctagt | ttttccggga | cgtaaaaata | taatggcagt | 240 |
| agctatcctt | tcctactac | cagcgaaacg | cgcgcgggtg | tttataattc | gagaacggat | 300 |
| aagccgtggc | cgggtggcgt | ttatgtgacg | cctgtgagca | gtgcgggggg | agtggcgatt | 360 |
| aaagctggct | cattaattgc | cgtgcttatt | ttgcgacaga | ccaacaacta | taacagcgat | 420 |
| gatttccagt | ttgtgtggaa | tatttacgcc | aataatgatg | tggtgggtgcc | cactggcggc | 480 |
| tgcgatgttt | ctgctcgtga | tgtcacccgt | actctgccgg | actaccctgg | ttcagtgccg | 540 |
| attcctctta | ccgtttattg | tgcgaaaagc | caaaacctgg | ggtattacct | ctccggcaca | 600 |
| accgcagatg | cgggcaactc | gattttcacc | aataccgcgt | cgttttcacc | cgcgcagggc | 660 |
| gtcggcgtag | agttggcgcg | caacggtagc | gttattccag | cgaataacac | ggtatcggtta | 720 |
| ggagcagtag | ggacttcggc | ggtgaagtctg | ggattaacgg | caaattacgc | acgtaccgga | 780 |
| gggcaggtga | ctgcagggaa | tgtgcaatcg | attattggcg | tgacttttgt | ttatcaa | 837 |

<210> 19
 <211> 837
 <212> DNA
 <213> E. coli

| | | | | | | |
|------------|------------|------------|------------|-------------|------------|-----|
| <400> 19 | | | | | | |
| ttcgcctgta | aaaccgccaa | tggtaccgca | atccctattg | gcggtggcag | cgccaatggt | 60 |
| tatgtaaacc | ttgcgcctgc | cgtgaatgtg | gggcaaaacc | tggtcgtaga | tctttcgacg | 120 |
| caaatctttt | gccataacga | ttaccagaa | accattacag | actatgtcac | actgcaacga | 180 |
| ggttcggcct | atggcgccgt | gttatctagt | ttttccggga | cgtaaaaata | taatggcagt | 240 |
| agctatcctt | tcctactac | cagcgaaacg | cgcgcgggtg | tttataattc | gagaacggat | 300 |
| aagccgtggc | cgggtggcgt | ttatgtgacg | ctggtgagca | gtgcgggggg | agtggcgatt | 360 |
| aaagctggct | cattaattgc | cgtgcttatt | ttgcgacaga | ccaacaacta | taacagcgat | 420 |
| gatttccagt | ttgtgtggaa | tatttacgcc | aataatgatg | tggtgggtgcc | cactggcggc | 480 |
| tgtgatgttt | ctgctcgtga | tgtcacccgt | actctgccgg | actaccctgg | ttcagtgccg | 540 |
| attcctctta | ccgtttattg | tgcgaaaagc | caaaacctgg | ggtattacct | atccggcaca | 600 |
| accgcagatg | cgggcaactc | gattttcacc | aataccgcgt | cgttttcacc | cgcgcagggc | 660 |

| | | | | | | |
|------------|------------|------------|------------|------------|------------|-----|
| gtcggcgtag | agttgacgcg | caacggtagc | attattccag | cgaataacac | ggtatcgta | 720 |
| ggagcagtag | ggacttcggc | ggtaagtctg | ggattaacgg | caaattacgc | acgtaccgga | 780 |
| gggcaggtga | ctgcagggaa | tgtgcaatcg | attattggcg | tgacttttgt | ttatcaa | 837 |

<210> 20
 <211> 837
 <212> DNA
 <213> E. coli

| | | | | | | |
|-------------|-------------|-------------|------------|-------------|------------|-----|
| <400> 20 | | | | | | |
| ttcgccgtgta | aaaccgccaa | tggtagcgct | atccctattg | gcggtaggcag | cgctaattgt | 60 |
| tatgtaaacc | ttgcgcctgc | cgtgaatgtg | gggcaaaacc | tggtcgtaga | tctttcgacg | 120 |
| caaactctttt | gccataacga | ttatccggaa | accattacag | actatgtcac | actgcaacga | 180 |
| ggctcggtctt | atggcggtgt | gttatctaata | ttttccggga | ccgtaaaata | tagtggcagt | 240 |
| agctatccat | ttccgaccac | cagcgaaacg | ccgcgggttg | tttataattc | gagaacggat | 300 |
| aagccgtggc | cggtggcgct | ttatttgacg | cctgtgagca | gtgcggggcg | ggtggcgatt | 360 |
| aaagctggct | cattaattgc | cgtgcttatt | ttgcgacaga | ccaaaaacta | taacagcgat | 420 |
| gattttccagt | ttgtgtggaa | tatttacgcc | aataatgatg | tggtagtgcc | tactggcggc | 480 |
| tgcatgtgtt | ctgctcgtga | tgtcaccgtt | actctgccgg | actaccctgg | ttcagtgcc | 540 |
| attcctctta | ccgtttattg | tgcgaaaagc | caaaacctgg | ggtattacct | ctccggcaca | 600 |
| accgcagatg | cggtgcaactc | gattttcacc | aataccgcgt | cgttttcacc | agcgaggggc | 660 |
| gtcggcgtag | agttgacgcg | caacggtagc | attattccag | cgaataacac | ggtatcgta | 720 |
| ggaacagtag | gaacttcggc | ggtaagtctg | ggattaacgg | caaattacgc | acgtaccgga | 780 |
| gggcaggtga | ctgcagggaa | tgtgcaatcg | attattggcg | tgacttttgt | ttatcaa | 837 |

<210> 21
 <211> 837
 <212> DNA
 <213> E. coli

| | | | | | | |
|-------------|-------------|-------------|------------|-------------|------------|-----|
| <400> 21 | | | | | | |
| ttcgccgtgta | aaaccgccaa | tggtagcgct | atccctattg | gcggtaggcag | cgccaatgtt | 60 |
| tatgtaaacc | ttgcgcctgc | cgtgaatgtg | gggcaaaacc | tggtcgtaga | tctttcgacg | 120 |
| caaactctttt | gccataacga | ttatccggaa | accattacag | actatgtcac | actgcaacga | 180 |
| ggctcggtctt | atggcggtgt | gttatctaata | ttttccggga | ccgtaaaata | tagtggcagt | 240 |
| agctatccat | ttccctaccac | cagcgaaacg | ccgcgcgttg | tttataattc | gagaacggat | 300 |
| aagccgtggc | cggtggcgct | ttatttgacg | cctgtgagca | gtgcggggcg | ggtggcgatt | 360 |
| aaagctggct | cattaattgc | cgtgcttatt | ttgcgacaga | ccaaacaacta | taacagcgat | 420 |
| gattttccagt | ttgtgtggaa | tatttacgcc | aataatgatg | tggtgggtgcc | tactggcggc | 480 |
| tgcatgtgtt | ctgctcgtga | tgtcaccgtt | actctgccgg | actaccctgg | ttcagtgcc | 540 |
| attcctctta | ccgtttattg | tgcgaaaagc | caaaacctgg | ggtattacct | ctccggcaca | 600 |
| accgcagatg | cggtgcaactc | gattttcacc | aataccgcgt | cgttttcacc | tgacaggggc | 660 |
| gtcggcgtag | agttgacgcg | caacggtagc | attattccag | cgaataacac | ggtatcgta | 720 |
| ggagcagtag | ggacttcggc | ggtagtctg | ggattaacgg | caaattatgc | acgtaccgga | 780 |
| gggcaggtga | ctgcagggaa | tgtgcaatcg | attattggcg | tgacttttgt | ttatcaa | 837 |

<210> 22
 <211> 837
 <212> DNA
 <213> E. coli

```

<400> 22
ttcgccctgta aaaccgccaa tgggtaccgca atccctattg gcggtggcag cgccaatggt 60
tatgtaaacc ttgcgcctgc cgtgaatgtg gggcaaaacc tggtcgtaga tctttcgacg 120
caaatctttt gccataacga ttaccagaa accattacag actatgtcac actgcaacga 180
ggtgcggtt atggcggtt gttatctagt tttccggga ccgtaaaata taatggcagt 240
agctatcctt tccctactac cagcgaaacg ccgcggttg ttataattc gagaacggat 300
aagccgtggc cgtggtgct ttatttgacg ccggtgagca gtgcggggg agtggtgatt 360
aaagctggct cattaattgc cgtgcttatt ttgcgacaga ccaacaacta taacagcgat 420
gatttcaggt ttgtgtggaa tatttacgcc aataatgatg tgggtggtgcc cactggcggc 480
tgcatgttt ctgctcgtga tgtcaccgtt actctgccgg actaccctgg ttcagtgcg 540
attcctctta ccgtttattg tgcgaaaagc caaacctgg ggtattacct ctccggcaca 600
accgcagatg cgggcaactc gattttcacc aataccgctg cgttttcacc cgcgcagggc 660
gtcggcgtag agttgacgca caacggtagc attattccag cgaataacac ggtatcgta 720
ggagcagtag ggacttcggc ggtaagtctg ggattaacgg caattacgc acgtaccgga 780
gggcaggtga ctgcagggaa tgtgcaatcg attattggcg tgacttttgt ttatcaa 837

```

```

<210> 23
<211> 279
<212> PRT
<213> E. coli

```

```

<400> 23
Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
1 5 10 15

Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val Asn Val Gly Gln
20 25 30

Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
35 40 45

Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
50 55 60

Gly Gly Val Leu Ser Asn Phe Ser Gly Ile Val Lys Tyr Ser Gly Ser
65 70 75 80

Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
85 90 95

Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
100 105 110

Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
115 120 125

Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
130 135 140

Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
145 150 155 160

Cys Asp Ala Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Arg

```

11

130 135 140
 Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
 145 150 155 160
 Cys Asp Ala Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Arg
 165 170 175
 Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
 180 185 190
 Leu Gly Tyr Tyr Leu Ser Gly Thr His Ala Asp Ala Gly Asn Ser Ile
 195 200 205
 Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
 210 215 220
 Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
 225 230 235 240
 Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
 245 250 255
 Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
 260 265 270
 Gly Val Thr Phe Val Tyr Gln
 275
 <210> 25
 <211> 279
 <212> PRT
 <213> E. coli
 <400> 25
 Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
 1 5 10 15
 Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly Gln
 20 25 30
 Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
 35 40 45
 Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ala Ala Tyr
 50 55 60
 Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val Lys Tyr Asn Gly Ser
 65 70 75 80
 Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
 85 90 95
 Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val

| 100 | 105 | 110 |
|--|-----|-----|
| Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val 115 120 125 | | |
| Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe 130 135 140 | | |
| Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly 145 150 155 160 | | |
| Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Arg 165 170 175 | | |
| Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn 180 185 190 | | |
| Leu Gly Tyr Tyr Leu Ser Gly Thr His Ala Asp Ala Gly Asn Ser Ile 195 200 205 | | |
| Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln 210 215 220 | | |
| Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu 225 230 235 240 | | |
| Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr 245 250 255 | | |
| Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile 260 265 270 | | |
| Gly Val Thr Phe Val Tyr Gln 275 | | |
| <210> 26 | | |
| <211> 279 | | |
| <212> PRT | | |
| <213> E. coli | | |
| <400> 26 | | |
| Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly 1 5 10 15 | | |
| Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly Gln 20 25 30 | | |
| Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr 35 40 45 | | |
| Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ala Ala Tyr 50 55 60 | | |
| Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val Lys Tyr Asn Gly Ser | | |

65 70 75 80
 Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
 85 90 95
 Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
 100 105 110
 Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
 115 120 125
 Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
 130 135 140
 Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
 145 150 155 160
 Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Arg
 165 170 175
 Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
 180 185 190
 Leu Gly Tyr Tyr Leu Ser Gly Thr His Ala Asp Ala Gly Asn Ser Ile
 195 200 205
 Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
 210 215 220
 Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
 225 230 235 240
 Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
 245 250 255
 Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
 260 265 270
 Gly Val Thr Phe Val Tyr Gln
 275

 <210> 27
 <211> 279
 <212> PRT
 <213> E. coli

 <400> 27
 Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
 1 5 10 15
 Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly Gln
 20 25 30
 Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr

35 40 45
 Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
 50 55 60
 Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly Ser
 65 70 75 80
 Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
 85 90 95
 Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
 100 105 110
 Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
 115 120 125
 Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
 130 135 140
 Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
 145 150 155 160
 Cys Asp Val Ser Ala His Asp Val Thr Val Thr Leu Pro Asp Tyr Arg
 165 170 175
 Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
 180 185 190
 Leu Gly Tyr Tyr Leu Ser Gly Thr His Ala Asp Ala Gly Asn Ser Ile
 195 200 205
 Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
 210 215 220
 Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
 225 230 235 240
 Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
 245 250 255
 Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
 260 265 270
 Gly Val Thr Phe Val Tyr Gln
 275
 <210> 28
 <211> 279
 <212> PRT
 <213> E. coli
 <400> 28
 Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly

| | | | |
|---|-----|-----|-----|
| 1 | 5 | 10 | 15 |
| Ser Ala Asn Val Tyr Val Asn Leu Ala Ile Ala Val Asn Val Gly Gln | 20 | 25 | 30 |
| Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr | 35 | 40 | 45 |
| Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr | 50 | 55 | 60 |
| Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly Ser | 65 | 70 | 75 |
| Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn | 85 | 90 | 95 |
| Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val | 100 | 105 | 110 |
| Ser Ser Ala Gly Gly Val Val Ile Lys Ala Gly Ser Leu Ile Ala Val | 115 | 120 | 125 |
| Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe | 130 | 135 | 140 |
| Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly | 145 | 150 | 155 |
| Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Arg | 165 | 170 | 175 |
| Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn | 180 | 185 | 190 |
| Leu Gly Tyr Tyr Leu Ser Gly Thr His Ala Asp Ala Gly Asn Ser Ile | 195 | 200 | 205 |
| Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln | 210 | 215 | 220 |
| Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu | 225 | 230 | 235 |
| Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr | 245 | 250 | 255 |
| Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile | 260 | 265 | 270 |
| Gly Ala Thr Phe Val Tyr Gln | 275 | | |

<210> 29

<211> 279
 <212> PRT
 <213> E. coli

<400> 29

Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
 1 5 10 15

Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val Asn Val Gly Gln
 20 25 30

Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
 35 40 45

Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
 50 55 60

Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly Ser
 65 70 75 80

Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
 85 90 95

Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
 100 105 110

Ser Ser Ala Gly Gly Leu Val Ile Lys Ala Gly Ser Leu Ile Ala Val
 115 120 125

Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
 130 135 140

Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
 145 150 155 160

Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Arg
 165 170 175

Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
 180 185 190

Leu Gly Tyr Tyr Leu Ser Gly Thr His Ala Asp Ala Gly Asn Ser Ile
 195 200 205

Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
 210 215 220

Leu Thr Arg Asn Gly Thr Ile Ile Pro Thr Asn Asn Thr Val Ser Leu
 225 230 235 240

Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
 245 250 255

Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile

260 265 270
 Gly Val Thr Phe Val Tyr Gln
 275
 <210> 30
 <211> 280
 <212> PRT
 <213> E. coli
 <400> 30
 Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
 1 5 10 15
 Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly Gln
 20 25 30
 Asn Leu Val Val Asp Leu Ser Thr Gln Thr Phe Cys His Asn Asp Tyr
 35 40 45
 Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
 50 55 60
 Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly Ser
 65 70 75 80
 Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
 85 90 95
 Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
 100 105 110
 Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
 115 120 125
 Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
 130 135 140
 Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
 145 150 155 160
 Cys Asp Val Ser Ala His Asp Val Thr Val Thr Leu Pro Asp Tyr Arg
 165 170 175
 Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
 180 185 190
 Leu Gly Tyr Tyr Leu Ser Gly Thr His Ala Asp Ala Gly Asn Ser Ile
 195 200 205
 Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
 210 215 220
 Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu

225 230 235 240
 Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
 245 250 255
 Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
 260 265 270
 Gly Val Thr Phe Val Tyr Gln Glx
 275 280

<210> 31
 <211> 279
 <212> PRT
 <213> E. coli

<400> 31
 Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
 1 5 10 15
 Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly Gln
 20 25 30
 Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
 35 40 45
 Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
 50 55 60
 Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val Lys Tyr Asn Gly Ser
 65 70 75 80
 Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
 85 90 95
 Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
 100 105 110
 Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
 115 120 125
 Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
 130 135 140
 Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
 145 150 155 160
 Cys Asp Ala Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Arg
 165 170 175
 Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
 180 185 190

Leu Gly Tyr Tyr Leu Ser Gly Thr His Ala Asp Ala Gly Asn Ser Ile
 195 200 205
 Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
 210 215 220
 Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
 225 230 235 240
 Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
 245 250 255
 Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
 260 265 270
 Gly Val Thr Phe Val Tyr Gln
 275
 <210> 32
 <211> 279
 <212> PRT
 <213> E. coli
 <400> 32
 Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
 1 5 10 15
 Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly Gln
 20 25 30
 Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
 35 40 45
 Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
 50 55 60
 Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly Ser
 65 70 75 80
 Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
 85 90 95
 Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
 100 105 110
 Ser Ser Ala Gly Gly Val Val Ile Lys Ala Gly Ser Leu Ile Ala Val
 115 120 125
 Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
 130 135 140
 Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
 145 150 155 160

Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Pro
 165 170 175
 Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
 180 185 190
 Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn Ser Ile
 195 200 205
 Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
 210 215 220
 Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
 225 230 235 240
 Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
 245 250 255
 Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
 260 265 270
 Gly Val Thr Phe Val Tyr Gln
 275

<210> 33
 <211> 279
 <212> PRT
 <213> E. coli

<400> 33
 Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
 1 5 10 15
 Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly Gln
 20 25 30
 Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
 35 40 45
 Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ala Ala Tyr
 50 55 60
 Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val Lys Tyr Asn Gly Ser
 65 70 75 80
 Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
 85 90 95
 Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
 100 105 110
 Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
 115 120 125

Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
 130 135 140
 Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
 145 150 155 160
 Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Pro
 165 170 175
 Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
 180 185 190
 Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn Ser Ile
 195 200 205
 Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
 210 215 220
 Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
 225 230 235 240
 Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
 245 250 255
 Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
 260 265 270
 Gly Val Thr Phe Val Tyr Gln
 275
 <210> 34
 <211> 279
 <212> PRT
 <213> E. coli
 <400> 34
 Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
 1 5 10 15
 Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val Asn Val Gly Gln
 20 25 30
 Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
 35 40 45
 Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
 50 55 60
 Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly Ser
 65 70 75 80
 Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
 85 90 95

Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
 100 105 110
 Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
 115 120 125
 Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
 130 135 140
 Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
 145 150 155 160
 Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Pro
 165 170 175
 Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
 180 185 190
 Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn Ser Ile
 195 200 205
 Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
 210 215 220
 Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
 225 230 235 240
 Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
 245 250 255
 Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
 260 265 270
 Gly Val Thr Phe Val Tyr Gln
 275
 <210> 35
 <211> 279
 <212> PRT
 <213> E. coli
 <400> 35
 Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
 1 5 10 15
 Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly Gln
 20 25 30
 Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
 35 40 45
 Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
 50 55 60

Gly Ser Val Leu Ser Ser Phe Ser Gly Thr Val Lys Tyr Asn Gly Ser
 65 70 75 80
 Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
 85 90 95
 Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
 100 105 110
 Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
 115 120 125
 Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
 130 135 140
 Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
 145 150 155 160
 Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Arg
 165 170 175
 Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
 180 185 190
 Leu Gly Tyr Tyr Leu Ser Gly Thr His Ala Asp Ala Gly Asn Ser Ile
 195 200 205
 Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
 210 215 220
 Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
 225 230 235 240
 Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
 245 250 255
 Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
 260 265 270
 Gly Val Thr Phe Val Tyr Gln
 275

<210> 36
 <211> 279
 <212> PRT
 <213> E. coli

<400> 36
 Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
 1 5 10 15
 Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val Asn Val Gly Gln
 20 25 30

Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
 35 40 45
 Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
 50 55 60
 Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly Ser
 65 70 75 80
 Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
 85 90 95
 Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
 100 105 110
 Ser Ser Ala Gly Gly Val Val Ile Lys Ala Gly Ser Leu Ile Ala Val
 115 120 125
 Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
 130 135 140
 Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
 145 150 155 160
 Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Arg
 165 170 175
 Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
 180 185 190
 Leu Gly Tyr Tyr Leu Ser Gly Thr His Ala Asp Ala Gly Asn Ser Ile
 195 200 205
 Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
 210 215 220
 Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
 225 230 235 240
 Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
 245 250 255
 Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Arg Ser Ile Ile
 260 265 270
 Ala Val Thr Phe Val Tyr Gln
 275

<210> 37
 <211> 279
 <212> PRT
 <213> E. coli

<400> 37

Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
 1 5 10 15
 Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly Gln
 20 25 30
 Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
 35 40 45
 Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
 50 55 60
 Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Glu Tyr Ser Gly Ser
 65 70 75 80
 Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
 85 90 95
 Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
 100 105 110
 Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
 115 120 125
 Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
 130 135 140
 Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
 145 150 155 160
 Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Arg
 165 170 175
 Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
 180 185 190
 Leu Gly Tyr Tyr Leu Ser Gly Thr His Ala Asp Ala Gly Asn Ser Ile
 195 200 205
 Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
 210 215 220
 Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
 225 230 235 240
 Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
 245 250 255
 Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
 260 265 270
 Gly Val Thr Phe Val Tyr Gln
 275

<210> 38
 <211> 279
 <212> PRT
 <213> E. coli

<400> 38

Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
 1 5 10 15
 Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly Gln
 20 25 30
 Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
 35 40 45
 Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
 50 55 60
 Gly Gly Val Leu Ser His Phe Ser Gly Thr Val Lys Tyr Ser Gly Ser
 65 70 75 80
 Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
 85 90 95
 Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
 100 105 110
 Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Met Ala Val
 115 120 125
 Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
 130 135 140
 Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
 145 150 155 160
 Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Arg
 165 170 175
 Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
 180 185 190
 Leu Gly Tyr Tyr Leu Ser Gly Thr His Ala Asp Ala Gly Asn Ser Ile
 195 200 205
 Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
 210 215 220
 Leu Thr Arg Asn Gly Thr Ile Asn Pro Ala Asn Asn Thr Val Ser Leu
 225 230 235 240
 Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
 245 250 255

Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
 260 265 270

Gly Val Thr Phe Val Tyr Gln
 275

<210> 39
 <211> 279
 <212> PRT
 <213> E. coli

<400> 39
 Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
 1 5 10 15

Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly Gln
 20 25 30

Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
 35 40 45

Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
 50 55 60

Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly Ser
 65 70 75 80

Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
 85 90 95

Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
 100 105 110

Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
 115 120 125

Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
 130 135 140

Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
 145 150 155 160

Cys Asp Val Ser Val Arg Asp Val Thr Val Ile Leu Pro Asp Tyr Arg
 165 170 175

Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
 180 185 190

Leu Gly Tyr Tyr Leu Ser Gly Thr His Ala Asp Ala Gly Asn Ser Ile
 195 200 205

Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
 210 215 220

Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
225 230 235 240

Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
245 250 255

Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Lys Ser Ile Ile
260 265 270

Gly Val Thr Phe Val Tyr Gln
275

<210> 40
<211> 279
<212> PRT
<213> E. coli

<400> 40
Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
1 5 10 15

Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly Gln
20 25 30

Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
35 40 45

Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
50 55 60

Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val Lys Tyr Asn Gly Ser
65 70 75 80

Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
85 90 95

Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
100 105 110

Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
115 120 125

Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
130 135 140

Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
145 150 155 160

Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Arg
165 170 175

Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
180 185 190

Leu Gly Tyr Tyr Leu Ser Gly Thr His Ala Asp Ala Gly Asn Ser Ile
 195 200 205
 Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
 210 215 220
 Leu Thr Ala Asn Gly Thr Ile Val Pro Ala Asn Asn Thr Val Ser Leu
 225 230 235 240
 Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
 245 250 255
 Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
 260 265 270
 Gly Val Thr Phe Val Tyr Gln
 275
 <210> 41
 <211> 279
 <212> PRT
 <213> E. coli
 <400> 41
 Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
 1 5 10 15
 Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly Gln
 20 25 30
 Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
 35 40 45
 Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
 50 55 60
 Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val Lys Tyr Asn Gly Ser
 65 70 75 80
 Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
 85 90 95
 Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Leu Val
 100 105 110
 Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
 115 120 125
 Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
 130 135 140
 Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
 145 150 155 160

Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Arg
 165 170 175
 Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
 180 185 190
 Leu Gly Tyr Tyr Leu Ser Gly Thr His Ala Asp Ala Gly Asn Ser Ile
 195 200 205
 Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
 210 215 220
 Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
 225 230 235 240
 Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
 245 250 255
 Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
 260 265 270
 Gly Val Thr Phe Val Tyr Gln
 275
 <210> 42
 <211> 279
 <212> PRT
 <213> E. coli
 <400> 42
 Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
 1 5 10 15
 Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val Asn Val Gly Gln
 20 25 30
 Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
 35 40 45
 Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
 50 55 60
 Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Asn Gly Ser
 65 70 75 80
 Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
 85 90 95
 Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
 100 105 110
 Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
 115 120 125

Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
 130 135 140
 Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
 145 150 155 160
 Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Arg
 165 170 175
 Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
 180 185 190
 Leu Gly Tyr Tyr Leu Ser Gly Thr His Ala Asp Ala Gly Asn Ser Ile
 195 200 205
 Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
 210 215 220
 Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
 225 230 235 240
 Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
 245 250 255
 Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
 260 265 270
 Gly Val Thr Phe Val Tyr Gln
 275
 <210> 43
 <211> 279
 <212> PRT
 <213> E. coli
 <400> 43
 Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
 1 5 10 15
 Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly Gln
 20 25 30
 Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
 35 40 45
 Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
 50 55 60
 Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly Ser
 65 70 75 80
 Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
 85 90 95

Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
 100 105 110
 Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
 115 120 125
 Leu Ile Leu Arg Gln Thr Lys Asn Tyr Asn Ser Asp Asp Phe Gln Phe
 130 135 140
 Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
 145 150 155 160
 Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Arg
 165 170 175
 Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
 180 185 190
 Leu Gly Tyr Tyr Leu Ser Gly Thr His Ala Asp Ala Gly Asn Ser Ile
 195 200 205
 Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
 210 215 220
 Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
 225 230 235 240
 Gly Thr Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
 245 250 255
 Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
 260 265 270
 Gly Val Thr Phe Val Tyr Gln
 275

<210> 44
 <211> 279
 <212> PRT
 <213> E. coli

<400> 44
 Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
 1 5 10 15
 Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val Asn Val Gly Gln
 20 25 30
 Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
 35 40 45
 Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
 50 55 60

Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly Ser
 65 70 75 80
 Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
 85 90 95
 Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
 100 105 110
 Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
 115 120 125
 Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
 130 135 140
 Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
 145 150 155 160
 Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Pro
 165 170 175
 Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
 180 185 190
 Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn Ser Ile
 195 200 205
 Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
 210 215 220
 Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
 225 230 235 240
 Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
 245 250 255
 Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
 260 265 270
 Gly Val Thr Phe Val Tyr Gln
 275

<210> 45
 <211> 279
 <212> PRT
 <213> E. coli

<400> 45
 Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
 1 5 10 15
 Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly Gln
 20 25 30

Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
 35 40 45
 Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ala Ala Tyr
 50 55 60
 Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val Lys Tyr Asn Gly Ser
 65 70 75 80
 Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
 85 90 95
 Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
 100 105 110
 Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
 115 120 125
 Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
 130 135 140
 Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
 145 150 155 160
 Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Arg
 165 170 175
 Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
 180 185 190
 Leu Gly Tyr Tyr Leu Ser Gly Thr His Ala Asp Ala Gly Asn Ser Ile
 195 200 205
 Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
 210 215 220
 Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
 225 230 235 240
 Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
 245 250 255
 Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
 260 265 270
 Gly Val Thr Phe Val Tyr Gln
 275

<210> 46
 <211> 7416
 <212> DNA
 <213> Artificial

<220>

<223> Sequence of Plasmid pCGA126-1

```

<400> 46
gcgcaccaata cgcaaacccgc ctctccccgc gcgttgccgc attcattaat gcagctggca 60
cgacagggttt cccgactgga aagcgggcag tgagcgcaac gcaattaatg tgagttagct 120
cactcattag gcaccccagg ctttacactt tatgcttccg gctcgtatgt tgtgtggaat 180
tgtgagcgga taacaatttc acacaggaaa cagctatgag taataaaaaac gtcaatgtaa 240
ggaaatcgca ggaaataaca ttctgcttgc tggcaggat cctgatgttc atggcaatga 300
tggttgccgg acgcgctgaa gcgggagtg ccttagtg cactcgcgta atttatccgg 360
cagggcaaaa acaagtgcaa cttgccgtga caaataatga tgaaaatagt acctatttaa 420
ttcaatcatg ggtggaaaat gccgatggtg taaaggatgg tcgttttatc gtgacgcctc 480
ctctgtttgc gatgaaggga aaaaaagaga atacctacg tattcttgat gcaacaaata 540
accaattgcc acaggaccgg gaaagtttat tctggatgaa cgtaaagcg attccgctca 600
tggataaatc aaaattgact gagaatacgc tacagctcgc aattatcagc cgcattaaac 660
tgtactatcg cccggctaaa tttagcgttg caccgatca ggccgcagaa aaattaagat 720
ttcgtcgtag cgcgaattct ctgacgctga ttaacccgac accctattac ctgacggtaa 780
cagagttgaa tgccggaacc cgggttcttg aaaatgcatt ggtgcctcca atgggcgaaa 840
gcacgggttaa attgccttct gatgcaggaa gcaatattac ttaccgaaca ataaatgatt 900
atggcgccact tacccccaaa atgacggcg taatggaata atagatctca cgacagggtt 960
ccgactgga aagcgggcag tgagcgcaac gcaattaatg tgagttagct cactcattag 1020
gcaccccagg ctttacactt tatgcttccg gctcgtatgt tgtgtggaat tgtgagcgga 1080
taacaatttc acacaggaaa cagctatgaa acgagttatt accctgtttg ctgtactgct 1140
gatgggctgg tcggtaaatg cctggtcatt cgcctgtaaa accgccaatg gtaccgctat 1200
ccctattggc ggtgcgacgc ccaatgttta tgtaaacctt gcgcccgctg tgaatgtggg 1260
gcaaaacctg gtcgtggatc ttctgacgca aatcttttgc cataacgatt atccggaaac 1320
cattacagac tatgtcacac tgcaacgagg ctccgcttat ggccgctgt tatctaattt 1380
ttccgggacc gtaaaatata gtggcagtag ctatccattt cctaccacca gcgaaacgcc 1440
gcgcgttgtt tataattcga gaacggataa gccgtggccg gtggcgcttt atttgacgcc 1500
tgtgagcagt gcggcgggg agctggctca ttaattgccg tgcttatttt 1560
gcgacagacc aacaactata acagcgatga ttccagttt gtgtggaata ttacgccaa 1620
taatgatgtg gtgtgccta ctggcgctg cgatgtttct gctcgtgatg tcaccgttac 1680
tctgccggac taccctggtt cagtgcacat tcctcttacc gtttattgtg cgaaaagcca 1740
aaacctgggg tattacctc cgggcacac cgcagatgcg ggcaactcga ttttcaccaa 1800
tacccgctg ttttcacctg cacaggcgct tggcgtagc acttcggcg tgagtctggg 1860
tattccagcg aataacacgg tatcgttagg agcagtaggg acttcggcg tgagtctggg 1920
atlaacggca aattatgcac gtaccggagg gcaggtgact gcagggaatg tgcaatcgat 1980
tattggcgtg acttttgttt atcaataagg atccgtcgac ctgcaggcat gcaagcttgg 2040
cactggccgt cgttttataa cgtcgtgact gggaaaaacc tggcgttacc caacttaatc 2100
gccttgacg acatccccct ttccgcagct ggcgtaatag cgaagaggcc cgcaccgatc 2160
gcccttccca acagttgcgc agcctgaatg gcgaatggcg cctgatgcgg tattttctcc 2220
ttacgcatct gtgcggtatt tcacaccgca taaattccct gttttggcgg atgagagaag 2280
attttcagcc tgatacagat taaatcagaa cgcagaagcg gtctgataaa acagaatttg 2340
cctggcgga gtagcgcggt ggtcccacct gaccccatgc cgaactcaag aagtgaacg 2400
ccgtagcgcc gatggtagt ttgggtctc cccatgcgag agtagggaac tgcaggcat 2460
caaataaaac gaaaggctca gtcgaaagac tgggccttct gttttatctg ttgtttgtcg 2520
gtgaacgctc tcctgagtag gacaaatccg ccgggagcgg atttgaacgt tgcgaagcaa 2580
cggcccgggg tggcggggca ggacgcccgc catatactgc caggcatcaa attaaagcata 2640
aggccatctg acgtatggcc tttttgcgtt tttaaacct cttccgtcca cctgacccca 2700
tgccgaactc aaagtgaaac gccgtagcgc cgatggtagt gtgggtctc cccatgcgag 2760
agtagggaac tgccaggcat caaataaaac gaaaggctca gtcgaaagac tgggcctttc 2820
gttttatctg ttgtttgtcg gtgaacgctc tcctgagtag gacaaatccg cccgggagcg 2880
gatttgaacg ttgcgaagca acggcccgga ggggtggcgg caggacgcc gccataaact 2940
gccaggcatc aaattaagca gaaggccatc ctgacggatg gcctttttgc gtttctacaa 3000
actcttctg tcgtcatatc tacaagccat cccccacag atacggtaaa ctacgctcgt 3060

```

| | | | | | | |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| ttttgcatca | ggaaagcagg | gaatttatgg | tgcactctca | gtacaatctg | ctctgatgcc | 3120 |
| gcatagttaa | gccagccccc | acacccgcca | acacccgctg | acgcgccctg | acgggcttgt | 3180 |
| ctgctcccg | catccgctta | cagacaagct | gtgaccgtct | ccgggagctg | catgtgtcag | 3240 |
| agggtttcac | cgctcatcacc | gaaacgcgcg | agacgaaagg | gccatgaaca | ataaaactgt | 3300 |
| ctgcttacat | aaacagtaat | acaaggggtg | ttatgagcca | tattcaacgg | gaaacgtctt | 3360 |
| ctctaggccg | cgattaaatt | ccaacatgga | tgtgtattta | tatgggtata | aatgggctcg | 3420 |
| cgataatgtc | gggcaatcag | gtgcgacaat | ctatcgattg | tatgggaagc | ccgatgcgcc | 3480 |
| agagttgttt | ctgaaacatg | gcacaggtag | cgctgccaat | gatgttacag | atgagatggt | 3540 |
| cagactaaac | tggctgacgg | aatttatgcc | tcttcgacca | tcaaccattt | tatccgtact | 3600 |
| cctgatgatg | catggttact | caccactgcg | atccccggaa | aacagcattc | caggtattag | 3660 |
| aagaatatcc | tgattcaggt | gaaaatatgt | ttgatgcgct | ggcagtggtc | ctgcgcgggt | 3720 |
| tgcattcgat | tectgtttgt | aattgtcctt | ttacacgcga | tcgcgtattt | cgtctcgctc | 3780 |
| aggcgcaatc | accaaagaat | aacggtttgg | ttgatgcgag | tgattttgat | gacgagcgta | 3840 |
| atggctggcc | tggtgaacaa | gtctggaaag | aaatgcataa | acttttgcca | ttctcaccgg | 3900 |
| attcagtcgt | cactcatggt | gattttctac | ttaataacct | tatttttgac | gaggggaaat | 3960 |
| taataggttg | tattgatgtt | ggacgagtcg | gaatcgcaga | ccgataccag | gatcttgcca | 4020 |
| tctatggaa | ctgcctcggt | gagttttctc | cttcattaca | gaaacggcct | tttcaaaaat | 4080 |
| atggtattga | taatcctgat | atgaataaat | tgcagtttca | tttgatgctc | gatgagtttt | 4140 |
| tctaagaatt | aattcatggc | cctcgtgata | cgctattttt | tataggttaa | tgcatgata | 4200 |
| ataatggttt | cttagacgtg | aggttctgta | cccgacacca | tcgaatggcg | caaaaccttt | 4260 |
| cgcggtatgg | catgatagcg | cccggaaagag | agtcaattca | gggtggtgaa | tgtgaaacca | 4320 |
| gtaacgttat | acgatgtcgc | agagtatgcc | ggtgtctctt | atcagaccgt | ttcccgctg | 4380 |
| gtgaaccagg | ccagccacgt | ttctgcgaaa | acgcgggaaa | aagtggaagc | ggcgatggcg | 4440 |
| gagctgaatt | acattcccaa | ccgcgtggca | caacaactgg | cgggcaaaaca | gtcgttgctg | 4500 |
| attggcggtg | ccacctccag | tctggccctg | cacgcgccgt | cgcaaatgtt | cgcggcgatt | 4560 |
| aaatctcgcg | ccgatcaact | gggtgcacgc | gtggtggtgt | cgatggtaga | acgaagcggc | 4620 |
| gtcgaagcct | gtaaagcggc | ggtgcacaa | cttctcgcgc | aacgcgtcag | tgggctgatc | 4680 |
| attaactatc | cgctggatga | ccaggatgcc | attcgttggt | aagctgcctg | cactaatggt | 4740 |
| ccggcggtat | ttcttgatgt | ctctgaccag | acacccatca | acagtattat | tttctcccat | 4800 |
| gaagacggtg | cgcgactggg | cgtggagcat | ctggtcgcac | tgggtcacca | gcaaatcgcg | 4860 |
| ctgttagcgg | gccattaaag | ttctgtctcg | gcgcgtctgc | gtctggctgg | ctggcataaa | 4920 |
| tatctcactc | gcaatcaaat | tcagccgata | gcggaacggg | aaggcgactg | gagtgccatg | 4980 |
| tccggttttc | aacaaaacct | gcaaatgctg | aatgagggca | tcgttccac | tgcgatgctg | 5040 |
| gttgccaacg | atcagatggc | gctgggcgca | atgcgcgcca | ttaccgagtc | cgggctgcgc | 5100 |
| gttggtgcgg | atatctcggt | agtgggatac | gacgataccg | aagacagctc | atgttatatc | 5160 |
| ccgcggttaa | ccaccatcaa | acaggatttt | cgctgtgctg | ggcaaacacc | cgtggaccgc | 5220 |
| ttgctgcaac | tctctcaggg | ccaggcggtg | aaggggcaatc | agctgttgcc | cgtctcactg | 5280 |
| gtgaaaaagaa | aaaccaccct | ggcgcccaat | acgcaaacccg | cctctccccg | cgcgttggcc | 5340 |
| gattcattaa | tgacgtggc | acgacaggtt | tcccgactgg | aaagcgggca | gtgagcgcaa | 5400 |
| cgcaattaat | ggttaagttg | ctcactcatt | aggcacccca | aggctttaca | ctttatgctt | 5460 |
| ccgacctgga | agaacctgac | gtcaggtggc | acttttcggg | gaaatgtgcg | cggaacccct | 5520 |
| atgtgtttat | ttttctaaat | acattcaaat | atgtatccgc | tcagtagaca | ataaccctga | 5580 |
| taaatgcttc | aataatattg | aaaaaggaag | agtatgagta | ttcaacattt | ccgtgtcgcc | 5640 |
| cttattccct | tttttgccgc | attttgcctt | cctgtttttg | ctcaccaga | aacgctgggtg | 5700 |
| aaagtaaaag | atgctgaaga | tcagttgggt | gcacgagtgg | gttacatcga | actggatctc | 5760 |
| aacagcggta | agatccttga | gagttttcgc | ccggaagaac | gttttccaat | gatgagcact | 5820 |
| tttaaagttc | tgctatgtgg | cgcggtatta | tcccgatttg | acgcccggca | agagcaactc | 5880 |
| gggtcgccgc | atacactatt | ctcagaatga | cttggttgag | tactcaccag | tcacagaaaa | 5940 |
| gcatcttacg | gatggcatga | cagtaagaga | attatgcagt | gctgccataa | ccatgagtga | 6000 |
| taacactgcg | gccaacttac | ttctgacaac | gatcgaggga | ccgaaggagc | taaccgcttt | 6060 |
| tttgacacaac | atgggggatc | atgtaactcg | ccttgatcgt | tgggaaccgg | agctgaatga | 6120 |
| agccatacca | aacgacgagc | gtgacaccac | gatgcctgta | gcaatgcaac | aacgttggcg | 6180 |
| caaaactatta | actggcgaaac | tacttactct | agcttcccg | caacaattaa | tagactggat | 6240 |
| ggaggcgcat | aaagttgcag | gaccacttct | gcgctcgccc | cttccggctg | gctggtttat | 6300 |

```

tgctgataaa tctggagccg gtgagcgtgg gtctcgcggg atcattgcag cactggggcc 6360
agatggtaag ccctcccgtg tcgtagttat ctacacgacg gggagtcagg caactatgga 6420
tgaacgaaat agacagatcg ctgagatagg tgcctcactg attaagcatt ggtaactgtc 6480
agaccaagtt tactcatata tacttttagat tgatttaaaa cttcattttt aattttaaag 6540
gatctagggtg aagatccctt ttgataatct catgaccaa atcccttaac gtgagtttcc 6600
gttccactga gcgtcagacc ccgtagaaaa gatcaaagga tcttcttgag atcctttttt 6660
tctgcgcgta atctgctgct tgcaaacaaa aaaaccaccg ctaccagcgg tggtttgttt 6720
gccggatcaa gagctaccaa ctctttttcc gaaggtaact ggcttcagca gacgcagat 6780
accaaatact gttcttctag tgtagccgta gctaggccac cacttcaaga actctgtage 6840
accgcctaca tacctcgtc tgctaactct gttaccagtg gctgctgcca gtggcgataa 6900
gtcgtgtctt accgggttgg actcaagacg atagttaccg gataaggcgc agcgtcggg 6960
ctgaacgggg ggttcgtgca cacagcccag cttggagcga acgacctaca ccgaactgag 7020
atacctacag cgtgagctat gagaaagcgc cacgcttccc gaaggagaa aggcggacag 7080
gtatccggta agcggcaggg tcggaacagg agagcgcacg agggagcttc cagggggaaa 7140
cgctgtgtat ctttatagtc ctgtcgggtt tcgccacctc tgacttgagc gtcgattttt 7200
gtgatgctcg tcaggggggc ggagcctatg gaaaaacgcc agcaacgcgg cctttttacg 7260
gttcctggcc ttttgctggc cttttgctca catgttcttt cctgcgttat cccctgatcc 7320
tgtggataac cgtattaccg cctttgagtg agctgatacc gctcgcgcga gccgaacgac 7380
cgagcgcagc gagtcagtga gcgaggaagc ggaaga 7416

```

<210> 47
 <211> 726
 <212> DNA
 <213> Artificial

<220>
 <223> Sequence of J96 fimC plus native signal sequence

```

<400> 47
atgagtaata aaaacgtcaa tgtaaggaaa tcgcaggaaa taacattctg cttgctggca 60
ggtatcctga tgttcatggc aatgatgggt gccggacgcg ctgaagcggg agtggcctta 120
ggtgcgactc gcgtaattta tccggcaggg caaaaacaag tgcaacttgc cgtgacaaat 180
aatgatgaaa atagtaccta tttaattcaa tcatgggtgg aaaatgccga tgggtgaaag 240
gatggtcggt ttatcgtgac gcctcctctg tttgcgatga agggaaaaaa agagaatacc 300
ttacgtatcc ttgatgcaac aaataaccaa ttgccacagg accgggaaag tttattctgg 360
atgaacgtta aagcgattcc gtcaatggat aaatcaaaat tgactgagaa tacgtacag 420
ctcgcaatta tcagccgcat taaactgtac tatcgcccg ctaaattagc gttgccaccc 480
gatcaggccg cagaaaaatt aagatttcgt cgtagcgcga attctctgac gctgattaac 540
ccgacaccct attacctgac ggtaacagag ttgaatgccg gaaccgggt tcttgaaaat 600
gcattggtgc ctccaatggg cgaaagcacg gttaaattgc cttctgatgc aggaagcaat 660
attacttacc gaacaataaa tgattatggc gcacttacc ccaaaatgac gggcgtaatg 720
gaataa 726

```

<210> 48
 <211> 903
 <212> DNA
 <213> Artificial

<220>
 <223> Sequence J96 fimH plus native signal sequence

<400> 48

| | | | | | | |
|------------|-------------|-------------|------------|------------|------------|-----|
| atgaaacgag | ttattaccct | gtttgctgta | ctgctgatgg | gctggtcggt | aaatgcctgg | 60 |
| tcattcgect | gtaaaaccgc | caatggtacc | gctatcccta | ttggcgggtg | cagcgccaat | 120 |
| gtttatgtaa | accttgcgcc | cgctgtgaat | gtggggcaaa | acctggtcgt | ggatctttcg | 180 |
| acgcaaatct | tttgccataa | cgattatccg | gaaaccatta | cagactatgt | cacactgcaa | 240 |
| cgaggctcgg | cttatggcgg | cgtgttatct | aatttttccg | ggaccgtaaa | atatagtggc | 300 |
| agtagctatc | catttcctac | caccagcgaa | acgccgcgcg | ttgtttataa | ttcgagaacg | 360 |
| gataagccgt | ggccgggtgg | gctttatgtg | acgcctgtga | gcagtgcggg | cggggtggcg | 420 |
| attaaagctg | gctcattaat | tgccgtgctt | attttgcgac | agaccaacaa | ctataacagc | 480 |
| gatgatttcc | agtttgtgtg | gaatattttac | gccaataatg | atgtggtggt | gcctactggc | 540 |
| ggctgcgatg | ttctgctcgt | tgatgtcacc | gttactctgc | cggactaccc | tggttcagtg | 600 |
| ccaattcctc | ttaccgttta | ttgtgcgaaa | agccaaaacc | tggggtatta | cctctccggc | 660 |
| acaaccgcag | atgcggggcaa | ctcgattttc | accaataccg | cgtcgttttc | acctgcacag | 720 |
| ggcgtcggcg | tacagttgac | gcgcaacggg | acgattatct | cagcgaataa | cacggtatcg | 780 |
| ttaggagcag | tagggacttc | ggcgggtgag | ctgggattaa | cggcaaatga | tgacgctacc | 840 |
| ggagggcagg | tgactgcagg | gaatgtgcaa | tcgattattg | gcgtgacttt | tgtttatcaa | 900 |
| taa | | | | | | 903 |

<210> 49
 <211> 814
 <212> DNA
 <213> Artificial

<220>
 <223> Sequence of kanamycin R gene

| | |
|------------|---|
| <400> 49 | |
| atgagccata | ttcaacggga aacgtcttgc tctaggccgc gattaaattc caacatggat 60 |
| gctgatttat | atgggtataa atgggctcgc gataatgtcg ggcaatcagg tgcgacaatc 120 |
| tatcgattgt | atgggaagcc cgatgcgcca gagttgttcc tgaacatgg cacaggtagc 180 |
| gctgccaatg | atgttacaga tgagatggtc agactaaact ggctgacgga atttatgcct 240 |
| cttcgaccat | caaccatttt atccgtactc ctgatgatgc atggttactc accactgcga 300 |
| tcccgcgaaa | acagcattcc aggtattaga agaatatcct gattcagggtg aaaatattgt 360 |
| tgatgcgctg | gcagtgttcc tgcgcgggtt gcattcgatt cctgtttgta attgtccttt 420 |
| taacagcgat | cgcgattttc gtctcgctca ggcgcaatca ccaatgaata acggtttggg 480 |
| tgatgcgagt | gattttgatg acgagcgtaa tggctggcct gttgaacaag tctggaaaga 540 |
| aatgcataaa | cttttgccat tctcaccgga ttcagtcgtc actcatggtg atttctcact 600 |
| taataacctt | atttttgacg aggggaaatt aataggttgt attgatgttg gacgagtcgg 660 |
| aatcgagac | cgataccagg atcttgccat cctatggaac tgctcgggtg agttttctcc 720 |
| ttcattacag | aaacgggttt ttcaaaaata tgggtattgat aatcctgata tgaataaatt 780 |
| gcagtttcat | ttgatgctcg atgagttttt ctaa 814 |

<210> 50
 <211> 1085
 <212> DNA
 <213> Artificial

<220>
 <223> Sequence of Lac IQ

| | |
|------------|--|
| <400> 50 | |
| atgtgaaacc | agtaacgtta tacgatgtcg cagagtatgc cgggtgtctct tatcagaccg 60 |
| tttcccgcgt | ggtaaccag gccagccacg tttctgcgaa aacgcgggaa aaagtggag 120 |

```

cgccgatggc ggagctgaat tacattccca acccgctggc acaacaactg gcgggcaaac 180
agtcgttgct gattggcggt gccacctcca gtctggccct gcacgcgccg tcgcaaattg 240
tcgcggcgat taaatctcgc gccgatcaac tgggtgccag cgtgggtggg tcgatggtag 300
aacgaagcgg cgtcgaagcc tgtaaagcgg cgggtgcaca tcttctcgcg caacgcgtca 360
gtgggctgat cattaactat ccgctggatg accaggatgc cattgctgtg gaagctgcct 420
gcactaatgt tccggcggtt tttcttgatg tctctgacca gacacccatc aacagtatta 480
ttttctccca tgaagacggg acgcgactgg cgggtggagca tctggctcgc ttgggtcacc 540
agcaaactgc gctgttagcg ggcccattaa gttctgtctc ggcgcgctct cgtctggctg 600
gctggcataa atatctcact cgcaatcaaa ttcagccgat agcggaacgg gaaggcgact 660
ggagtgccat gtccgggttt caacaaacca tgcaaatgct gaatgagggc atcgttccca 720
ctgcgatgct ggttgccaac gatcagatgg cgtggggcgc aatgcgcgcc attaccgagt 780
ccgggctgcg cgttggtgcg gatctctcgg tagtgggata cgacgatacc gaagacagct 840
catgttatac cccgccgtta accaccatca aacaggattt tcgcctgctg gggcaaacca 900
gcgtggaccg cttgctgcaa ctctctcagg gccaggcggg gaagggaact cagctgttgc 960
ccgtctcact ggtgaaaaga aaaaccaccc tggcgcccaa tacgcaaac gcctctcccc 1020
cgcggttggc cgattcatta atgcagctgg cagcacaggg ttcccgactg gaaagcgggc 1080
agtdga 1085

```

<210> 51
 <211> 862
 <212> DNA
 <213> Sequence of beta-lacyamase gene

```

<400> 51
atgagtatte aacattttccg tgtcgccctt attccctttt ttgcggcatt ttgccttcct 60
gtttttgtct acccagaaac gctgggtgaa gtaaaagatg ctgaagatca gttgggtgca 120
cgagtgggtt acatcgaact ggatctcaac agcggtaaga tccttgagag ttttcgcccc 180
gaagaacgtt ttccaatgat gagcactttt aaagtcttgc tatgtggcgc ggtattatcc 240
cgtattgacg ccgggcaaga gcaactcggg tcgcccata cactattctc agaatgactt 300
ggttgagtac tcaccagtca cagaaaagca tcttacggat ggcatgacag taagagaatt 360
atgcagtgtc gccataacca tgagtgataa cactgcggcc aacttacttc tgacaacgat 420
cggaggaccg aaggagctaa ccgctttttt gcacaacatg ggggatcatg taactcgctt 480
tgatcgttg gaaacggagc tgaatgaagc cataccaaac gacgagcgtg acaccacgat 540
gcctgtagca atgcaacaac gttggcgcaa actattaact ggcgaaactac ttactctagc 600
ttcccgga caattaatag actggatgga ggccgataaa gttgcaggac cacttctgcg 660
ctcgccctt ccggtggct ggtttattgc tgataaatct ggagccgggt agcgtgggtc 720
tcgcggtatc attgcagcac tggggccaga tggtaagccc tcccgatcgt tagttatcta 780
cacgacgggg agtcaggcaa ctatggatga acgaaataga cagatcgctg agataggtgc 840
ctcactgatt aagcattggg aa 862

```

<210> 52
 <211> 601
 <212> DNA
 <213> Artificial

<220>
 <223> Sequence of the origin of replication

```

<400> 52
tttttctgcy cgtaatctgc tgcttgcaaa caaaaaaac accgctacca gcggtggttt 60
gtttgccgga tcaagagcta ccaactcttt ttccgaagg aactggcttc agcagagcgc 120
agataccaaa tactgttctt ctagtgtagc cgtagctagg ccaccacttc aagaactctg 180

```

```

tagcaccgcc tacataacctc gctctgctaa tcctgttacc agtggtgct gccagtggcg 240
ataagtcgtg tcttaccggg ttggactcaa gacgatagtt accggataag gcgcagcggg 300
cgggctgaac ggggggttcg tgcacacagc ccagcttggg gcgaacgacc tacaccgaac 360
tgagatacct acagcgtgag ctatgagaaa gcgccacgct tcccgaaggg agaaaggcgg 420
acaggatatcc ggtaagcggc aggggtcgaa caggagagcg cacgaggag cttccagggg 480
gaaacgcctg gtatctttat agtctgtcg ggtttcgcca cctctgactt gagcgtcgat 540
ttttgtgatg ctgctcaggg gggcggagcc tatggaaaaa cgccagcaac gcggcctttt 600
t 601

```

<210> 53
 <211> 116
 <212> DNA
 <213> Sequence of Lac p/o

```

<400> 53
cgcaattaat gtgagttagc tcactcatta ggcacccag gctttacact ttatgettcc 60
ggctcgtatg ttgtgtggaa ttgtgagcgg ataacaattt cacacaggaa acagct 116

```

<210> 54
 <211> 837
 <212> DNA
 <213> E. coli

```

<400> 54
ttcgccctgta aaaccgccaa tgggtaccgt atccctattg gcggtggcag cgccaatgtt 60
tatgtaaaacc ttgcgcccggt cgtgaatgtg gggcaaaaacc tggtcgtgga tctttcgacg 120
caaatctttt gccataacga ttatccggaa accattacag actatgtcac actgcaacga 180
ggctcggctt atggcggcgt gttatctaatt tttccggga ccgtaaaata taatggcagt 240
agctatccat ttctaccac cagcgaaacg ccgcgcgttg ttataattc gagaacggat 300
aagccgtggc cgggtggcgt ttatttgacg cctgtgagca gtgcgggcgg ggtggcgatt 360
aaagctggct cattaattgc cgtgcttatt ttgcgacaga ccaacaacta taacagcgat 420
gatttccagt ttgtgtggaa tatttacgcc aataatgatg tgggtgtgcc tactggcggc 480
tgcgatgtt ctgctcgtga tgtcacggtt actctgccgg actaccctgg ttcagtgcc 540
attcctctta ccgtttattg tgcgaaaagc caaacctgg ggtattacct ctccggcaca 600
accgcagatg cgggcaactc gattttcacc aataccgcgt cgttttcacc tgcacagggc 660
gtcggcgtag agttgacgcg caacggtacg attattccag cgaataacac ggtatcgta 720
ggagcagtag ggacttcggc ggtgagtcgt ggattaacgg caaattatgc acgtaccgga 780
gggcaggtga ctgcaggga tgtgcaatcg attattggcg tgacttttgt ttatcaa 837

```

<210> 55
 <211> 279
 <212> PRT
 <213> Artificial

<220>
 <223> Consensus sequence of FimH proteins for SEQ ID NO: 23 to 45

<400> 55
 Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly

| | | | |
|---|-----|-----|-----|
| 1 | 5 | 10 | 15 |
| Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly Gln | 20 | 25 | 30 |
| Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr | 35 | 40 | 45 |
| Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr | 50 | 55 | 60 |
| Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly Ser | 65 | 70 | 75 |
| Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn | 85 | 90 | 95 |
| Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val | 100 | 105 | 110 |
| Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val | 115 | 120 | 125 |
| Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe | 130 | 135 | 140 |
| Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly | 145 | 150 | 155 |
| Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Arg | 165 | 170 | 175 |
| Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn | 180 | 185 | 190 |
| Leu Gly Tyr Tyr Leu Ser Gly Thr His Ala Asp Ala Gly Asn Ser Ile | 195 | 200 | 205 |
| Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln | 210 | 215 | 220 |
| Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu | 225 | 230 | 235 |
| Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr | 245 | 250 | 255 |
| Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile | 260 | 265 | 270 |
| Gly Val Thr Phe Val Tyr Gln | 275 | | |

<210> 56
<211> 55
<212> DNA
<213> Artificial

<220>

<223> Oligonucleotide primer GA1F

<400> 56
cctgccatgg cgggtgtggc gctgggtgcg acccgcgta tttatccggc agggc 55

<210> 57
<211> 36
<212> DNA
<213> Artificial

<220>

<223> Oligonucleotide primer GA1R

<400> 57
ggcgtcgaca gattctatta ttccattacg cccgtc 36

<210> 58
<211> 36
<212> DNA
<213> Artificial

<220>

<223> Oligonucleotide primer GA13F

<400> 58
cacacaggaa acagctatga ttgtaatgaa aacgag 36

<210> 59
<211> 39
<212> DNA
<213> Artificial

<220>

<223> Oligonucleotide primer GA6R

<400> 59
ggcgtcgacg gatccttatt gataaacaaa agtcacgcc 39

<210> 60
<211> 30
<212> DNA
<213> Artificial
<220>
<223> Oligonucleotide primer GA11F

<400> 60
ccgaataaag atatcacgac aggtttcccg 30

<210> 61
<211> 19
<212> DNA
<213> Artificial

<220>
<223> Oligonucleotide primer GA9R

<400> 61
catagctgtt tcctgtgtg 19

<210> 62
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Oligonucleotide primer GA24F

<400> 62
tgctcacatg ttctttcctg cgt 23

<210> 63
<211> 34
<212> DNA
<213> Artificial

<220>
<223> Oligonucleotide primer GA23R

<400> 63
gacgttttta ttactcatag ctgtttcctg tgtg 34

<210> 64
<211> 37
<212> DNA
<213> Artificial

<220>
<223> Oligonucleotide primer GA21F

<400> 64
atgagtaata aaaacgtcaa tgtaaggaaa tcgcagg 37

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
17 January 2002 (17.01.2002)

PCT

(10) International Publication Number
WO 02/004496 A3

(51) International Patent Classification: C12N 15/31,
C07K 14/245, 16/12, C12N 15/72, A61K 39/108, 39/40

(21) International Application Number: PCT/US01/21525

(22) International Filing Date: 6 July 2001 (06.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/216,750 7 July 2000 (07.07.2000) US

(71) Applicant: MEDIMMUNE, INC. [US/US]; 35 West
Watkins Mill Road, Gaithersburg, MD 20878 (US).

(72) Inventors: LANGERMANN, Solomon; 6606 Cross
Country Boulevard, Baltimore, MD 20878 (US). REVEL,
Andrew; 5106 West Mockingbird Lane, Dallas, TX 75209
(US). AUGUSTE, Christine; 20147 Laurel Hill Way,
Germantown, MD 20874 (US). BURLEIN, Jeanne; 5313
Juxon Place, Springfield, VA 22151 (US).

(74) Agents: GRANT, Alan, J. et al.; Carella, Byrne, Bain,
Gilfillan, Cecchi, Stewart & Olstein, 6 Becker Farm Road,
Roseland, NJ 07068 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments

(88) Date of publication of the international search report:
25 July 2002

*For two-letter codes and other abbreviations, refer to the "Guide-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

WO 02/004496 A3

(54) Title: FIMH ADHESIN PROTEINS AND METHODS OF USE

(57) Abstract: The present invention provides bacterial immunogenic agents for administration to humans and non-human animals to stimulate an immune response. Also provided are methods for vaccination of mammalian species, especially human patients, with variants of the *E. coli* FimH protein, said variants being derived from different strains of *E. coli*, and to production of antibodies that protect the vaccine recipient against infection by pathogenic bacterial species. In another aspect the invention provides antibodies against such proteins and protein complexes that may be used as diagnostics and/or as protective/treatment agents for pathogenic bacterial species. A plasmid-based method of producing polypeptides, especially fused polypeptides, such as the complex of a bacterial chaperone and a bacterial adhesin, is also disclosed.

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 01/21525

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/31 C07K14/245 C07K16/12 C12N15/72 A61K39/108
A61K39/40

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EP0-Internal, WPI Data, PAJ, BIOSIS, CHEM ABS Data, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|--|-----------------------------|
| X | WO 95 20657 A (GX BIOSYSTEMS AS ;SOKURENKO EVGENI VENIAMINOVIC (US); HASTY DAVID) 3 August 1995 (1995-08-03) SEQ ID NO:1 page 43, paragraph 2; table 1 page 4, line 4 - line 8 --- | 1-3,7,8, 10, 13-30,32 |
| P,X | WO 01 04148 A (MEDIMMUNE INC) 18 January 2001 (2001-01-18) SEQ ID NO:1 page 8, line 14 - line 24 --- -/-- | 1-3,7,8, 10, 13-30,32 |

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"8" document member of the same patent family

Date of the actual completion of the international search

25 February 2002

Date of mailing of the international search report

22 05. 2002

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Mata-Vicente, M

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/21525

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|---|---|
| A | <p>LANGERMANN S ET AL: "VACCINATION WITH FIMH ADHESIN PROTECTS CYNOMOLGUS MONKEYS FROM COLONIZATION AND INFECTION BY UROPATHOGENIC ESCHERICHIA COLI" JOURNAL OF INFECTIOUS DISEASES, CHICAGO, IL, US, vol. 181, February 2000 (2000-02), pages 774-778, XP000971593 ISSN: 0022-1899 page 777, paragraph 3 -----</p> | <p>1-3,7,8, 10, 13-30, 32,34, 36-38</p> |

Best Available Copy

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 01/21525

| Patent document cited in search report | | Publication date | Patent family member(s) | Publication date |
|---|---|---------------------|----------------------------|---------------------|
| WO 9520657 | A | 03-08-1995 | AU 1532795 A | 15-08-1995 |
| | | | CA 2180726 A1 | 03-08-1995 |
| | | | WO 9520657 A1 | 03-08-1995 |
| | | | EP 0738325 A1 | 23-10-1996 |
| ----- | | | | |
| WO 0104148 | A | 18-01-2001 | AU 6094000 A | 30-01-2001 |
| | | | EP 1194561 A2 | 10-04-2002 |
| | | | WO 0104148 A2 | 18-01-2001 |
| ----- | | | | |